Essays in the Economics of Child Health and Skill Formation

Giacomo Mason

A dissertation submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

University College London (UCL)

Department of Economics

June 3, 2019

I, Giacomo Mason, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Recent research on human capital development during childhood has focused on three important avenues, among others: measurement, modelling, and interventions. In this thesis, in touch on each of these in turn. The chapter titled *"The effect of cash and information on child development"* examines the child development effects of a "cash plus" intervention in Nigeria, which starts from the pregnancy period. It underlines the interplay between resources and information in achieving growth and cognition improvements. Chapter *"Inequality in socioemotional skills"* highlights issues of measurement. It finds that there is no perfect invariance in the measurement of socioemotional skills in two cohorts of British five year olds born 30 years apart, and shows that socioeconomic determinants of such skills have changed over this period. Finally, the chapter titled *"The role of diet quality and physical activity in the production of adolescent human capital"* models the human capital production process in early adolescence, exploiting novel sources of exogenous variation to disentangle the health effects of diet and exercise. Significant complementarities between physical and mental health, and between mental health and diet, emerge from the analysis.

Impact statement

The second chapter of this thesis, titled *“The effect of cash and information on child development"*, evaluates the Child Development Grant Programme (CDGP), a combined cash transfer and information intervention in Northern Nigeria to tackle child underdevelopment. The rigorous quantitive evaluation of aid interventions via randomised controlled trials (RCTs) is increasingly used as a tool by policymakers in the developed and developing world. This type of investigation aims to answer the question “Does X work to solve issue Y?". The randomised way in which “X" is assigned – in our case, we assign villages to receive the intervention – ensures that the estimated effects are not merely associations but instead are causal in nature.

Results from the evaluation of CDGP have already achieved a direct policy impact. In fact, they have been considered by the Federal Government of Nigeria in designing its new National Social Protection Policy. Secondly, our results contribute to the global knowledge base around aid effectiveness. The peculiar design of the intervention, combining cash and information, is part of a new wave of “cash plus" interventions. Ours is one of the first investigations into a “cash plus" programme operating at scale in a developing country. The results highlight a small but significant effect on child linear growth, which has often failed to materialise in pure unconditional cash transfers. We provide evidence that interventions combining information and cash have the potential to affect child outcomes, even in extremely difficult and deprived contexts such as Northern Nigeria. Finally, the data from our study will be made freely available to academic and policy researchers a year after the intervention has concluded. This will enable further research into the social and biological mechanisms underlying child development in poor areas of the world.

The chapter titled *“Inequality in socioemotional skills"* has profound implication for researchers using psychometric scales to assess child development. It highlights the importance of measurement invariance issues. Many studies in economics use scales to quantify skill levels for children belonging to different groups (gender, ethnicity, cohorts). We show that it is not a given that the measurement properties of the scale are the same across groups. In our application, we reject full measurement invariance for samples of UK children born 30 years apart, so that conclusions about secular trends in socioemotional skills cannot be straightforwardly drawn.

The final chapter of this thesis – *“The role of diet quality and physical activity in the production of adolescent human capital"* – touches on obesity and mental health in early adolescence. Both issues have been increasing in magnitude in the past decades, and more attention has been devoted to them in the public debate. I model their joint evolution between ages 11 and 14, and how diet quality and physical activity affect them. The main message is that it is important to consider these dimensions jointly and not in isolation, given that they are interrelated in complex ways.

Acknowledgements

Many people have contributed significantly to my growth across the PhD journey. First and foremost, I'm grateful to my supervisors Orazio Attanasio and Gabriella Conti for their guidance, challenge, support, and inspiration. Together with Sir Richard Blundell, they have also been delightful coauthors on the third chapter of this thesis.

The experience working on the evaluation of the Child Development Grant Programme is something I would not trade with anything else. I thank the Institute for Fiscal Studies, and especially Pedro Carneiro and Imran Rasul for hiring me on this project and for being such stimulating coauthors. The opportunity to travel to Nigeria and gain field experience has been invaluable, on both a professional and personal level. For this I am grateful to Oxford Policy Management, and in particular to Lucie Moore and Molly Scott. I am also grateful to the whole staff at the OPM Abuja office, and the fearless squad of enumerators for CDGP – nothing would have been possible without them. Thanks also to ITAD and the rest of the ePact consortium, and to Save the Children and Action Against Hunger for implementing the programme in a difficult region.

Data is fundamental to answer policy questions. The work of the Centre for Longitudinal Studies in maintaining and expanding the British Cohorts is of disproportionate social value, and I am grateful to them for supporting me. In particular, Aida Sanchez and David Church have been instrumental in making secure access data for the final chapter accessible.

Despite not featuring in this thesis, my work as part of the DynaHEALTH consortium was very formative, and opened my eyes to the potential of multidisciplinary international research to address the greatest issues facing our societies.

I gratefully acknowledge the funding sources that made my Ph.D. work possible, namely the Economic and Social Research Council scholarship programme.

This journey would not have been possible were it not for my companions in the UCL Economics PhD programme. Interacting with such a diverse, brilliant, and humorous group was a personal highlight throughout these years. In particular, the freedom and irreverence of Drayton G06 has been an irreplaceable island of comfort.

I thank my family and friends back home in Italy for always supporting my pursuits, and for tolerating my ineptitude at long-distance communications as if I had never left. Finally, I'm grateful to Jenny for being by my side for most of this time, despite the distance and my frequent pedantic self-absorption.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The relevance of the childhood period is increasingly evidenced in the social and biological sciences. Most outcomes across the life course have their origin in development processes taking place early in life. Understanding the nature of this process is of paramount importance for the design of policies promoting efficiency and tackling inequalities in opportunity. This thesis is a collection of three essays touching on the most important topics in human capital research in the past decade: measurement, modelling, and interventions. The essays embrace different periods of childhood, from the womb to adolescence, and draw on data from both the developing and the developed world.

The recent literature in the economics of childhood human capital has advanced on three main fronts. Firstly, the issue of *measurement* has taken centre stage. It's now recognised that human capital is a multifaceted entity, encompassing cognition, socioemotional skills, physical and mental health, and more. Consequently, there is no unique, superior, or direct way in which human capital can be measured. Similarly, investment choices by parents into their own offspring are often not observed. However, concerted multidisciplinary efforts from the fields of statistics, economics, psychology, and education have equipped researchers with a wide range of measurement tools – such as scales and test batteries. Data collected with different tools can be viewed as imperfect but useful measures of lower-dimensional unobservable constructs. Recent powerful identification results show how these sources of information can be combined using factor analytic approach to efficiently extract the information on such latent constructs, while accounting for measurement error in what is observed.

Secondly, econometric techniques for *modelling* the process of human capital development have become salient. Different facets of human capital and investment are interrelated, and exhibit complex, dynamic patterns of complementarity and self-productivity in determining child development. Accordingly, it has become necessary to formulate production functions for human capital that are able to identify these features in a manner that is parsimonious enough to allow estimation. Finally, the potential of *interventions* to shed light on the process of human development has become clear. Evidence from interventions is especially useful when they are evaluated in an experimental framework, which can provide exogenous variation in relevant dimensions of human capital or investments and allow the identification of policy relevant aspects.

Chapter 2 of this thesis speaks directly to the latter point. In joint work with Pedro Carneiro, Imran Rasul, and Lucie Moore, we present evidence from the cluster randomised evaluation of the Child Development Grant Programme (CDGP), a large scale intervention in an extremely deprived part of Northern Nigeria. The CDGP provides pregnant women in rural and semi-rural areas with a monthly cash grant, until their child turns two. In parallel, it offers a behaviour change component to encourage correct pregnancy and child feeding practices. The rich survey we administer to households allows us to assess a plethora of effects, from labour supply and wealth to child nutrition and development. We find that CDGP led to small but significant increases in height and communication skills for children born after the intervention started. These effects are mostly explained by improvements in knowledge and practices around child feeding, and less by pure resource effects, pointing to a relevant separate effect of the informational component of the intervention. This work constitutes the first time the effects of a cash transfer at scale on child cognitive development is evaluated in Sub-Saharan Africa, and significantly advances the evidence on “cash plus" interventions in the developing world.

The issue of measurement is the main concern of chapter 3. In collaboration with Orazio Attanasio, Richard Blundell, and Gabriella Conti, we investigate the evolution of socioemotional skills in two cohorts of five year old children born 30 years apart in the UK. We focus on latent externalising and internalising skills, defined as misconduct/hyperactive behaviours and absence of emotional/somatic symptoms, respectively. The main challenge derives from the fact that two different scales are available for the two cohorts, hence the need to derive a novel sub-scale made up of eleven overlapping items. Still, there are many reasons why the two latent constructs might not be straightforwardly comparable across time – including different wording of the items, interpretation by parent respondents, and changing social norms. We thus set up a measurement invariance analysis through a multiple group confirmatory factor analysis model, which allows to empirically test the degree of comparability across cohorts. We find that children born in 1975 and 2005 can be assessed on the same scale, but individual levels on the latent dimensions are not comparable. In addition, the early-life determinants of socioemotional skills have changed in this 30-year arc, while they remain predictive of adolescent risky behaviours in a way that is independent of cognition. This work is important in underlining that the measurement properties of commonly used psychometric scales should not be taken for granted, as is often a-critically done in applied research.

Adolescence is a relatively understudied period in this literature. Chapter 4 examines the production of physical and mental health in this later period of childhood. In particular, it deals with the health implications of investments (namely, diet and physical activity) in a longitudinal sample of English and Welsh children, starting around age 11. Discovering the nature of the health production process in this period, which is still sensitive for later-life outcomes, can inform policy decisions around diet and physical activity in the school age. Not much evidence is currently available on complementarities between exercise and diet at this age. Here, the issue of modelling is central. I adopt a flexible functional form for the process of human capital production, that takes previous levels of health and current levels of investment as inputs. This specification is able to concisely capture many features of the process including complementarity and substitutability patterns. I rely on novel data on weather and food prices to disentangle the effect of diet and exercise. These sources of exogenous variation are used to instrument investments, following a control functions approach. I find that previous levels of health are fairly substitutable with exercise and diet in determining current health. Strong complementarities however emerge between mental and physical health. Better diet quality is significantly conducive to mental health, in particular for children that start from a poor diet.

Chapter 2

The effect of cash and information on child development: The Child Development Grant Programme in Northern Nigeria

2.1 Introduction

Children in Sub Saharan Africa (SSA) are disproportionately born into poverty. Despite recent advances, over 20% of under fives will be below the 1.90$ per day poverty line by 2030 (Watkins and Quattri, 2016). Poverty in the region intertwines with malnutrition. An estimated 33% of children under five are stunted, indicating long-term deficits in nutrition (Akombi et al., 2017). Northwestern Nigeria, the setting for this paper, fares even worse than the SSA average: child mortality sits at 185 per 1000 live births, and more than half of children under five are stunted (NPC and ICF, 2014).

Deprivation in early life has important consequences for human capital accumulation (Case et al., 2005). It hinders cognitive development and schooling achievement during childhood, and economic productivity in adulthood (Grantham-McGregor et al., 2007).[1] In the aggregate, deficits in human capital contribute to the intergenerational transmission of poverty. A recent study estimates a 7% per capita GDP penalty deriving from child stunting (Galasso and Wagstaff, 2018). At the same time, early interventions have the potential to reduce developmental deficits and improve long term outcomes.[2]

[1] Useful reviews of the long term consequences of early life stunting can be found in Dewey and Begum (2011) and McGovern et al. (2017).

[2] Long run effects have often been observed in developed contexts. For example, US programmes such as Perry (Anderson, 2008; Heckman et al., 2010, 2013) and Abecedarian (Campbell et al., 2014; Conti et al., 2016) led to improved education, employment, earnings, and health. But examples exist from developing contexts. For example, the Jamaican study shows that psychosocial stimulation in early life benefits educational and social outcomes in adolescence and adulthood (Grantham-McGregor et al., 1991; Walker et al., 2005, 2011). In Guatemala, improving early nutrition resulted in better education and productivity in adulthood (Hoddinott et al., 2008; Maluccio et al., 2009).

Effective design of early interventions relies on knowledge of the complex process of human capital development. Child human capital is multifaceted, encompassing domains such as linear growth, cognition, language, motor skills, and socio-emotional skills. Different domains can be affected by different factors – e.g. sanitation, diet, healthcare, psychosocial stimulation, and availability of resources. Furthermore, the development process exhibits sensitive (or critical) periods. During these periods, investments in human capital yield maximal benefits (Wachs et al., 2014). Sensitive periods differ across domains, depending on the biological processes underlying development.[3]

A widespread view has developed, which establishes the sensitive period for stunting to be early in life. Between birth and 24 months, deprived children accumulate a height deficit with respect to the healthy reference population. The deficit persists through childhood, with limited scope for catch-up growth. This phenomenon is known as growth faltering. Researchers and policymakers have pinpointed the first thousand days of life as critical 'window of opportunity' for nutrition (Victora et al., 2010). This period includes pregnancy, as maternal health and nutrition influence growth starting from the womb (Dewey and Begum, 2011).

While useful, the focus on the first thousand days has limitations. Prentice et al. (2013) observe that the cross-sectional evidence in Victora et al. (2010) should not be overinterpreted, and that catch-up growth at later ages is observed in many longitudinal datasets. As an example, Hirvonen (2014) finds that puberty might constitute a second window of opportunity to remediate earlier height disadvantage. More recently, methodological improvements in how to define and model catch-up growth have provided a more nuanced characterisation of the phenomenon. Anand et al. (2018) examine evidence from multiple developing countries, using a unified latent growth framework that allows to disentangle between-group (relative to a healthy population) and within-group dynamics. They find that the nature, extent, and velocity of catch-up growth is significantly different across contexts. Perumal et al. (2018) highlight the limitations of the use of stunting as an individual-level classifier of malnutrition.

Within the window of the the first thousand days, recent research has emphasised the role of protein in the introduction of complementary foods, particularly protein of animal origin. In particular, intake of animal source foods appears fundamental (Dewey and Adu-Afarwuah, 2008; Ghosh, 2016). Biomedical research has highlighted the link between essential amino acids (which cannot be synthesised by the human body, and need to be introduced with food) and the regulation and promotion of growth (Semba et al., 2016). At the population level, strong associations exist between consumption of animal protein in the first thousand days of life and child growth (Headey et al., 2017). Recent work in the human capital production function literature finds a significant role for protein intake in the development process (Puentes et al., 2016).

[3] A range of interventions to improve child development have been proposed and tested. The relative effectiveness of different types of intervention is compared in various reviews; see for example Bhutta et al., 2013; Black and Dewey, 2014; Grantham-McGregor et al., 2014; Fernald et al., 2017. A relatively recent literature focuses on characterising the process of skill formation, viewing it as a production function and explicitly modelling the process of parents' investment into their offspring's development. For different approaches to the issue see for example Cunha et al. (2010); Agostinelli and Wiswall (2016); Attanasio et al. (2017).

This paper investigates the impact of an early childhood development intervention in Northern Nigeria, the Child Development Grant Programme (CDGP). The intervention provides beneficiary women with cash and information. It starts during pregnancy, and lasts until their child turns two years of age. Our primary focus lies in the linear growth, communication, and motor skills of children born during the intervention.

This study innovates the literature in multiple important directions. We focus on an extremely deprived rural context in Sub Saharan Africa, for which less evidence exists. Our unique 'cash plus' intervention bundles a sizeable labelled unconditional cash transfer with an explicit educational component. Unlike most other studies in this area, the intervention starts during pregnancy and covers the first thousand days. Our findings have a causal interpretation, thanks to a clear experimental design. We measure multiple aspects of child development (linear growth, communication, and motor skills) for multiple children, born both before and after the intervention had been introduced. Our deep questionnaire contains a range of potential intermediate outcomes, such as parental knowledge, antenatal care, breastfeeding and complementary feeding, child diet, expenditures, and livelihoods. Finally, although this paper is limited to a two-year followup, we will be able to observe outcomes in the long term in a four-year horizon.

The results in this paper are intent-to-treat estimates based on a randomised trial clustered at the village level. We estimate CDGP increased the height-for-age of children born during the intervention by around a fifth of a standard deviation. At the same time, treated children are thinner for their height, and more likely to be wasted. Part of the effects on physical growth are due to treated children being born to longer pregnancies, and thus being relatively younger at follow-up. We observe a small effect on communication skills, and no effect on motor skills. Development of older siblings living in the same households was not affected. We then show that CDGP improved a range of potential intermediate inputs to child development. The design of our intervention does not allow to separate the role of cash and information. Even so, increased household resources can't fully explain many of the observed changes in intermediate inputs. This points to a separate role for information. A simple mediation analysis shows that the development effects are explained to a large extent by knowledge and practices, and much less by household resources themselves. Finally, we use Engel curves to highlight how the intervention changed food consumption patterns in a way that is incompatible with a simple increase in purchasing power.

Our work speaks to a range of research questions in the literature. Firstly, are cash transfers an effective tool for improving nutrition and development? In the past couple of decades, cash transfers have been widely employed with the aim of enhancing recipients' livelihoods. Cash can partially relieve households of liquidity constraints. It can thus allow households to increase investment in its members' human capital. This can take various forms: increased food intake and diversity, smoother food consumption across periods of cyclical or unanticipated scarcity, better access to healthcare services, etc.

Cash transfer interventions are heterogeneous. Among other aspects, they vary in terms of context, design, timing, duration, size, and conditionality. Reviews in both Gaarder et al.

(2010) and Manley et al. (2013) show overall ambiguous effects of cash transfers on children's nutritional status. This is despite the fact that transfers often improve access to healthcare and increase the amount and/or diversity of food consumed within the households.[4] A few studies in developing countries also present effects on other domains of child human capital. Paxson and Schady (2010) find that a UCT in Ecuador did not affect child development. Macours et al. (2012) find a CCT in Nicaragua had positive effects on vocabulary, language, memory, and socioemotional skills. To the best of our knowledge, ours is the first large-scale evaluation of a cash transfer programme in Sub Saharan Africa to assess cognitive development and motor skills.

Secondly, are cash transfers enough? Or does cash need other components to be effective? In particular, we speak to the literature on the effects of information and education interventions. Poor knowledge of best practices related to pregnancy and infant feeding might prevent deprived households from pursuing efficient choices. This might limit the usefulness of other components like cash transfers. Information-only interventions have shown promising results, increasing child height and weight (Bhutta et al., 2013).[5] These considerations have recently sparked interest in interventions that combine cash and information. More generally, there's been a shift towards 'cash-plus' interventions, integrating cash transfers with other types of complementary support (Roelen et al., 2017). To our knowledge, (Levere et al., 2016) is the only study where the role of cash and information can be neatly isolated.[6] Set in rural Nepal, their experiment has one arm receiving only an educational intervention. Another arm receives the same information together with cash. The authors find that the most sizeable improvements occur in the 'cash plus' arm. This is true even for outcomes that are not a priori connected with liquidity constraints, like parental knowledge and practices. This can indicate the presence of complementarities between the cash and information components.

In some sense, the inclusion of an information component alongside cash transfers is a well established practice. Conditionalities often tie the receipt of cash to participation in educational activities. In this perspective, they serve as an implicit 'plus' component.[7] Even

[4] A systematic review of the entire body of evidence on cash transfers is beyond the scope of this paper. Interested readers can refer to a number of insightful reviews, among which: Dewey and Adu-Afarwuah (2008); Leroy et al. (2009); Gaarder et al. (2010); Manley et al. (2013); Bastagli et al. (2016).

[5] Many of these interventions are small very intensive, and might be hard to implement at scale – see for example Penny et al. (2005) for Peru, Roy et al. (2005, 2007) for Bangladesh, Zaman et al. (2008) for Pakistan, and Zhang et al. (2013) for China. Notable exceptions are Linnemayr and Alderman (2011), who find weight improvements in a large-scale intervention in Senegal, and Fitzsimons et al. (2016), who estimate a large effect on height-for-age following a home visiting programme in Malawi. Reviews on the effectiveness of educational interventions are available e.g. in Imdad et al. (2011) for breastfeeding promotion, and Lassi et al. (2013) for promotion of complementary feeding in the 6-24 months period. A new wave of interventions under the umbrella of *Alive&Thrive* have been launched in various countries. They combine interpersonal counseling, community mobilisation, and mass media, and are implemented at scale, making them very similar to the Behaviour Change Component of CDGP. Results from Vietnam, Ethiopia, and Bangladesh show improvements in breastfeeding and complementary feeding practices, but not in child nutritional status (Menon et al., 2016; Kim et al., 2016; Frongillo et al., 2017; Kim et al., 2018).

[6] The ongoing Transfer Modality Research Initiative (`https://clinicaltrials.gov/ct2/show/NCT02237144`), implemented in Bangladesh by IFPRI, has a suitable experimental design to inform about the separate role of cash and information, but to our knowledge no results on nutrition and development have been published to date.

[7] Many of the conditional cash transfers in Central and South America require children to undergo preventive

unconditional transfers might have features that implicitly convey information about correct practices. For example, the cash might be labelled towards certain uses, or accompanied by explicit messaging around best practices. These features are known as ‘soft conditionalities’, to distinguish them from their ‘hard’ counterparts (Pace and Pellerano, 2016). The intervention evaluated in our paper combines messaging and labelling in a novel way for the literature.

A third contribution is around the existence of a critical period for stunting in early life. Our questionnaire allows us to observe growth, health, and dietary diversity for both children born before and after the intervention started. Despite similar improvements in diet and health outcomes, the older children do not see any change in their height or weight at midline, when aged around five. This indicates that intervening later, after birth, is less effective than intervening prenatally – seemingly providing further evidence for the existence of a very early critical window, possibly tied to breastfeeding and complementary feeding practices.

Fourthly, our paper also contributes to the growing body of evidence that underlines the importance of protein intake for linear growth. The food recall module in our questionnaire enables us to accurately measure dietary diversity, including different types of animal source foods. We observe large increases in the intake of animal protein, particularly meat and dairy, which can be a plausible mechanism explaining our height results.

Finally, this work touches upon the role of men in the adoption of health-promoting behaviours within households. Women, and mothers in particular, put more weight on the well-being of children. Targeting transfers to women can shift decision-making power in their favour (Jayachandran, 2015). In Björkman Nyqvist and Jayachandran (2017), the same educational sessions yield smaller improvements if provided to husbands rather than wives, even in a context where decision-making power is skewed towards males. In our case, CDGP gives transfers to mothers. However, the educational component is offered to both men and women, and we measure both parents’ knowledge.

The paper is structured as follows. Section 2.2 explores in detail the features of the CDGP programme. Section 2.3 describes the experiment and the data, and documents the programme’s coverage. Section 2.4 presents effects on child development, and contextualises them in the literature. Section 2.5 examines potential mechanisms driving the effects on child development, and provides evidence that the education component had an effect separate from the cash transfer. Section 2.6 concludes.

2.2 The CDGP intervention

The Child Development Grant Programme (CDGP) is a five-year pilot cash transfer and information programme targeting pregnant women in northern Nigeria. The main objective of the programme is to improve early life nutrition in rural and semi-rural areas that exhibit high

medical checkups and health monitoring, where mothers receive information and advice about practices conducive to their children’s nutrition and development. See for example Fernald et al. (2008) for *Oportunidades* in Mexico, Attanasio et al. (2005) for Colombia, and Macours et al. (2012) for Nicaragua. It’s not always possible to isolate the role of conditionalities; a notable exception is Attanasio et al. (2015), where the authors show that preventive visits promoted child health separately from the cash component.

rates of child malnourishment. The intervention has two components: (i) it provides pregnant women with an *unconditional cash transfer* – to tackle economic roots of inadequate nutrition such as poverty and food insecurity; (ii) and it offers an *educational component* – to address inadequate knowledge about appropriate practices around pregnancy and the feeding of infants and young children.

The CDGP was implemented in five Local Government Authorities (LGAs) across two states in Northern Nigeria.[8] These areas are almost exclusively rural, and subsistence farming is the main livelihood for many households (Solivetti, 1994). Families, predominantly of Hausa ethnicity and Muslim religion, are structured around a male household head, who is responsible for providing for the household, and one or more wives and their children. Women are mostly secluded in the household's walled compound, but can still engage in income-generating activities (e.g. food preparation, weaving) and can retain such income for their own priorities (Munro et al., 2011). Child nutrition practices are generally inadequate, with very low levels of early breastfeeding initiation, exclusive breastfeeding, and appropriate complementary feeding; this is accompanied by staggering levels of malnutrition, with early-life stunting rates between 55 and 60 percent (NPC and ICF, 2014).

The CDGP is implemented by Save the Children and Action Against Hunger. It builds on Save the Children's experience in developing and implementing nutritional interventions in the region.[9] A formative research process was undertaken to incorporate the views of stakeholders across the intervention areas, including households with young children, traditional and religious leaders, and health workers. The programme was subsequently trialed between April and July 2014 in fifteen pilot villages across both states, which were excluded from the evaluation.

2.2.1 Cash transfer

The resource component of the CDGP is an unconditional monthly cash transfer targeted to pregnant women. The grant was originally set at 3,500 Nigerian Naira (NGN) – 21.6 U.S. Dollars (USD) at the PPP exchange rate observed at inception of the programme on 15 August 2014.[10] It amounts to approximately 17% of mean total monthly household pre-programme expenditure and around 18% of the sum of monthly family earnings.[11] The transfer is even more substantial from the perspective of the women in our sample: it represents 130% of their mean pre-programme earnings from work activities. Women become eligible for the grant once pregnant. The monthly payments continue until the child turns two years old. Thus a

[8] The LGAs in question are Anka and Tsafe (in Zamfara state), and Buji, Gagarawa and Kiri Kasama (in Jigawa state). The programme is implemented by an international Non-Governmental Organisation (NGO) consortium. In Zamfara, the implementation is carried out by Save the Children (SC), while in Jigawa, the programme is being coordinated by Action Against Hunger (ACF). The programme is funded by the UK Department for International Development (DfID), and independently evaluated by a separate consortium, e-Pact, led by Oxford Policy Management (OPM) and Itad in collaboration with the Institute of Fiscal Studies (IFS). More information is available at `https://www.opml.co.uk/projects/evaluation-child-development-grant-programme-cdgp`.

[9] For example, the Working to Improve Nutrition in Northern Nigeria (WINNN) project (Visram et al., 2017)

[10] This figure was revised upwards to 4,000 NGN during the spring of 2016, to account for rising inflation.

[11] We define monthly family earnings as the sum of the woman's and her husband's earnings.

woman can potentially receive up to 33 payments (9 during pregnancy, and 24 after the birth of the child).

Rollout of CDGP happens on a village-by-village basis. After a short period of sensitisation and mobilisation of the community involving local and religious leaders (approximately a week), the programme teams start the process of targeting and enrolment. Potential beneficiary women are identified, and their eligibility is ascertained. To be eligible, women need to fulfil two requirements: (i) residing in a village where CDGP is operating, and (ii) being pregnant.[12] No other conditionalities are imposed for the receipt of the cash grant. Once eligibility is confirmed, women are officially enrolled in an electronic database for the payment system. They are provided with a mobile phone and a recharge card required to activate it; the number of the simcard acts as a unique woman ID for all programme-related activities.[13] Once a month, each village is visited by payment agents. They confirm the identity of beneficiary women using a thumbprint scanner, and provide the monthly payment directly to the women, in cash. [14]

2.2.2 Information component

The second core element of the CDGP is the Behaviour Change Communication (BCC) strategy. It is formulated around eight 'key messages' regarding health, nutrition, and infant and young child feeding (IYCF), as summarised in Panel A of Table 2.1. The key messages concern multiple stages of child development: similarly to the cash transfer, they start from pregnancy and end when the child turns two years old.[15]

The key messages are communicated through different channels, reported in Panel B of Table 2.1. This was intended to maximise the likelihood that residents in programme areas are exposed to the key messages, including people that are not directly targeted by the programme but are still influential in community life – such as husbands, traditional leaders, and elderly women (Sharp et al., 2016).

The Behaviour Change Communication channels can be grouped into two types. *Low-Intensity* channels – posters, radio, preaching, health talks, food demonstrations, and voice

[12] Pregnancy status is confirmed by a urine test. Various solutions were piloted for confirming pregnancy. The most accurate is testing by a trained health worker at a clinic, but it was deemed infeasible due to the poor coverage in health infrastructure – especially in Zamfara. Applicants were initially asked to provide a sample of early morning urine, but this approach proved easily subject to fraud. Pregnancy was ultimately tested on-the-spot, in the presence of a female Community Volunteer (Sharp et al., 2016).

[13] It was originally intended for the phones to enable access to mobile financial services, but this was revealed as infeasible. The phones are now mostly used as IDs, to alert beneficiaries about payment dates, and to disseminate part of the educational component. Thumbprints are also electronically recorded for identification purposes.

[14] If the woman dies after the child's birth, the payments are still disbursed to a female caregiver until the child turns two. There are some other cases in which a woman can exit the programme. In the case of miscarriage or stillbirth, the woman receives payments up to the month next to the event, and can re-enrol for a subsequent pregnancy. The same happens if the child dies, but the woman is not allowed to re-enrol. The woman also loses eligibility if she moves to a non-programme community, or in fraud cases (e.g. false urine test, false residency, double registration).

[15] The content of these messages was developed in accordance with the National Strategic Plan of Action for Nutrition (Federal Ministry of Health, 2014), and are based on the UNICEF Community Infant and Young Child Feeding Counselling Package (adapted for the Nigerian context). Some example instructional materials are pictured in Figure 2.A1.

messages – conform to a 'one-size-fits-all' approach to communication. Beneficiaries are passive recipients of messages, which are disseminated in multiple ways as to maximise exposure.[16] *High-Intensity* channels focus on more customised and interactive forms of communication and training, where beneficiaries participate actively. This intensity distinction serves as the basis for the differentiation of the treatment groups in our evaluation – see Section 2.3 below.

The Behaviour Change Communication strategy is implemented by Community Volunteers (CVs). These women are recruited locally within communities to receive a three-day training around the programme content. They are then deployed in their communities with the task of coordinating implementation and promoting the practices recommended by the Behaviour Change Communication curriculum.[17] Namely, they are responsible for the High-Intensity Behaviour Change Communication activities, i.e. infant and young child feeding support groups and one-to-one counselling. Support groups are collective training sessions taking place approximately once a month. A number of beneficiaries (ideally 12-15) meet with one or more Community Volunteers to receive training and information regarding infant and young child feeding practices, and discuss their experiences. Support groups are offered to beneficiaries' husbands as well, although attendance is lower. Separately from these meetings, beneficiaries can seek out one-to-one counselling sessions with a Community Volunteer in their own homes to deal with any specific issue. For issues beyond the scope of the Community Volunteer, mothers are referred directly to the health facility or hospital in the area.

On-the-ground implementation of the programme, especially in its Behaviour Change Communication component, initially witnessed substantial variation both in terms of timing and geographical coverage. This was mainly due to staffing and procurement issues in the initial teething phase. This is particularly true of some subcomponents. Health talks, food demonstrations, and infant and young child feeding support groups suffered delays in Jigawa state, starting only in April/May 2015 – i.e. about 6 months after the rollout of payments. The frequency and coverage of these activities also varied in time and between the two states (Sharp et al., 2016).

In sum, the cash component of CDGP provides pregnant women in deprived rural communities with a substantial amount of cash on a monthly basis. Unlike many other interventions that begin after birth (Paxson and Schady, 2010; Macours et al., 2012; Handa et al., 2016), the CDGP cash transfer is set up to be disbursed pre-conceptionally. The cash transfer is sizeable, amounting to almost 20% of average family income. Its disbursement continues for more than two years, providing households with a sustained increase in resources. The transfer is not subject to any conditionality, but it is implicitly 'labelled' towards the pregnancy and

[16]For some channels, no interaction is possible at all (e.g. posters, radio, voice messages); food demonstrations and health talks – dubbed Action Oriented Groups (AOGs) – are 'performed' live in the villages, but they are usually attended by a large number of people, thus preventing any significant level of interaction.

[17]Two types of CVs were recruited: *lead* Community Volunteers, who are more experienced, receive more intensive training, and cover specialised counselling roles; and *nutrition promotion* Community Volunteers, usually two per group, who receive more basic training and are tasked with promoting the recommended practices but refer to lead Community Volunteers when necessary. They do not receive any stipend for their time, but can expense transport and meals to CDGP.

the resulting child. The amount is nontrivial, especially given the relative low level of economic activity among women in programme areas. In parallel to the cash grant, the educational component of CDGP covers a wide range of topics, involving all phases of child development from conception to age two. The key behaviour change messages were delivered using a multitude of different channels of communication, with the aim of maximising exposure. The Behaviour Change Communication is designed to have two different levels of intensity, depending on the level of interactivity offered to recipients.

2.3 Evaluation design and data collection

2.3.1 The experiment

We conducted a cluster randomized controlled trial (RCT) to evaluate the impact of CDGP on household and child-level outcomes. The evaluation covered a total of 208 villages in the five Local Government Authorities where the programme is being rolled out. The unit of randomisation was the village.[18] The villages were randomly allocated to one of three arms:

- **Control (C)** group, where the programme does not operate – neither the cash transfer nor the educational component – until after the evaluation period has ended in 2018.[19]
- **Low-Intensity Information (T1)** group, where the cash transfer was offered, and the Behaviour Change Communication messages were distributed *only* via Low-Intensity channels – see Table 2.1.
- **High-Intensity Information (T2)** group, where the cash transfer was offered, and *both* Low- and High-Intensity Behaviour Change Communication channels were used.

This experimental set up enables us to rigorously and causally measure the impact of the CDGP and to study differential effects of the CDGP intervention, depending on the intensity of the Behaviour Change Communication component: we are able to test whether offering interactive and customised forms of communication is more effective than the more basic channels only. We are not however able to tease out the effect of cash separately from information, as in Levere et al. (2016) for example.

2.3.2 Sampling

We sampled 26 households in each village that contained either at least one pregnant woman or a woman likely to become pregnant based on socioeconomic characteristics. We first sampled one traditional ward per village because villages are on average too large to be surveyed

[18]The village was chosen as the unit of randomisation because boundaries between villages are well-defined. This is meant to minimise spillovers of programme components to control areas and reduce the possibilities of disputes. The evaluation selected the 208 villages from a list of all villages in the five LGAs, supplied by the programme implementers. Villages that were part of the CDGP piloting activities were excluded. In two instances, smaller villages had to be joined into the same primary sampling unit (PSU) because of their extreme proximity.

[19]The radio component of the Behaviour Change Communication strategy was for obvious reason not directly excludable.

entirely.[20] We then conducted a census of all households in each sampled traditional ward in order to identify households with at least one pregnant woman or a woman likely to become pregnant. In each village, all households with at least one pregnant woman were selected for the baseline survey. Once the list of pregnant women in the ward was exhausted, additional women were selected based on their probability of becoming pregnant in the subsequent two years, until the target cluster size of 26 households was met.

The likelihood of becoming pregnant was established using a prediction model based on data from the 2013 Nigeria Demographic and Health Survey (NPC and ICF, 2014). The probability of giving birth in the next two years was modelled as a function of woman's age, time since last birth, household size, number of children aged under and over 5 years in household, and TV ownership. The estimated coefficients from a linear probability model on the DHS data were then used to predict pregnancy probability in the CDGP listing data. Estimated pregnancy probabilities for non-pregnant women are visualised in Figure 2.A3. The women with the highest estimated probability were then included in the baseline sample if needed to reach the target of 26 women per village. Predicted probabilities for sampled women lie mostly above .5 (the mean and median of the distribution are .6 and .65 respectively), with no differences between treatment arms. A detailed illustration of the sampling procedure can be found in Appendix Section 2.9.1.

One 'index' woman was interviewed in each household. This results in two main subsamples:

- Households where the woman is pregnant at inception of the programme, and is therefore both *immediately eligible* to receive the cash transfer, and not subject to any immediate fertility response;
- Households where the woman is not pregnant at inception, but likely to become pregnant during the study, and may respond to the offer of cash by seeking to have additional children or by narrowing the spacing after her last birth.

A timeline of the evaluation process is depicted in Figure 2.A2. Sampled households were administered the baseline questionnaire shortly after the listing phase, from late August to October 2014. Immediately afterwards, the villages were randomised according to the experimental design, and programme rollout was started. Approximately two years later (October/November 2016), the same households were administered a midline questionnaire.[21]

[20] For the purpose of this survey, a household is defined as any group of people who live in the same dwelling unit and have common cooking and eating arrangements. A traditional ward is a further subdivision of a village community, usually made up of a separate cluster of households. The 'traditional' denomination is to distinguish it from the administrative ward, which is not superimposable. If the sampled traditional ward was too small (defined as containing less than 200 households in total), a neighbouring traditional ward was also sampled. If the sampled traditional ward was too large (defined as containing more than 200 households in total), the traditional ward was divided into equal parts and only one part was listed.

[21] Randomisation was carried out in three tranches, in order to minimise the delay between data collection and the actual implementation of the programme. This is accounted for in the estimation of treatment effects described below. A final round of data collection is planned for late 2018, just after the programme will have wrapped up. In parallel to the quantitative evaluation – household, community, and market surveys – two additional workstreams are taking place: a qualitative evaluation (Sharp and Cornelius, 2015; Sharp et al., 2018) and a process evaluation (Sharp et al., 2016).

It's important to stress that the sampling strategy employed in the data collection process does *not* yield a representative sample of households residing in evaluation areas, since only households with pregnant (or likely to become pregnant) women are selected. Moreover, the sampling strategy overweights households from smaller villages, since the target cluster size is the same independently on the size of the village. Reweighting the analysis to make it representative of a wider population would require an extensive census of villages in the areas where CDGP is implemented. The census carried out before the baseline interview cannot serve this purpose, as it is limited to the evaluation communities, which are predominantly rural and semi-rural. No attempt is thus made at constructing sample weights to make estimates representative of all households with pregnant (or likely to become pregnant) women residing in the sampled LGAs.

2.3.3 Data

This paper uses data from the baseline and midline surveys. We collected data about the 'index' woman, who is potentially eligible to receive the cash transfer, her household, her husband, and her children (if any). The woman is interviewed about her work activities, her knowledge about pregnancy and child feeding practices, the household's food availability, her participation to the cash transfer, and her exposure to the Behaviour Change Communication messaging. Her husband – if available to be interviewed – is administered similar modules. Household-level information about finances, assets, and expenditure is also collected.[22]

At baseline, one of the children of the woman aged 0-5 (if any are available) is randomly selected to be surveyed in detail; this will be referred to as *'old' child* from now onwards. In addition, if the woman gave birth to any offspring *after* the baseline survey (and thus after programme inception), this *'new' child* is administered the same questionnaire as the 'old' child at baseline. If more than one child has been born after baseline, a single new child at random. With this strategy, we can observe outcomes for children exposed to the intervention after birth (the 'old' children) as well as for children exposed starting from the pregnancy period (the 'new' children).

The main outcomes of interest are measures of child development. We collect information of children's growth and nutritional status by measuring their height, weight, and middle upper arm circumference. To ensure quality of the data, anthropometric data was collected by a dedicated enumerator who received separate specialised training. These are used to derive age-normed indicators of child development and nutritional status (WHO, 2009).

We additionally measure child development by administering two modules – communication and gross motor – from the Ages and Stages Questionnaire (ASQ-3, Squires and Bricker, 2009). The ASQ is a screening questionnaire that assesses a child's development by asking his/her caretaker whether the child is able to perform a number of specific tasks. Six

[22] If can happen that husbands are not in the household when the survey teams perform the interviews, since they are often working or cultivating land during the day. In such cases, if a revisit to the household is not possible, the household-level information is collected from another household member who is knowledgeable about these dimensions.

age-specific tasks are investigated for each of various sub-domains. These are different for children grouped in 2-3 month windows.[23] The original implementation of the ASQ covers children from birth to 60 months. The version of ASQ used in the CDGP survey – which was translated into Hausa language, adapted to the local context, and carefully piloted on the field – is instead limited to the range 5-37 months. The ASQ instrument has been meaningfully employed in a variety of developed and developing country contexts (Fernald et al., 2012; Levere et al., 2016; Doyle, 2017).[24]

For the remainder of the paper, we focus on results from a specific subset of households. Of the 5,433 women (residing in as many households) interviewed at baseline, we first select only those who reported being pregnant at baseline (3,688, 68%). Of these, 354 were not followed up at midline because the security situation in their areas had deteriorated. A further 109 households were not found at midline, refused consent, or relocated outside the evaluation areas. Overall, we have midline data for 3,225 household where the woman reported being pregnant at baseline.

We further restrict the sample to only those household in which a new child was found and surveyed at midline (2,718). The discrepancy with the sample of 3,225 is likely due to a combination of misreported pregnancies, unsuccessful pregnancies, and infant mortality. Finally, we only consider those children who, based on their reported date of birth, were estimated to be in utero at the time of the baseline survey – just before the programme started.[25] This gives a final sample of 2,216 households – which we call 'in-utero' sample – in which the woman was carrying a child at baseline who was still alive at midline. This sample therefore excludes any women whose decision to have a baby may have been influenced by the presence of CDGP. The resulting sample of 'in utero' new children are between 14 and 27 months old. Most of these households (73%) also contain an 'old child' who was alive at baseline. In this way, we are able to compare effects on children exposed to CDGP at different ages.

Characteristics of households at baseline do not differ substantially between treatment arms. Sample balance is assessed in Table 2.2 for the main estimation sample. Households in T1 display slightly higher per-capita expenditure than other groups, and women in T2 are more likely to be in a polygamous marriage; we account for this by adding these baseline covariates to the set of controls used for estimation.

Attrition during the study period affected around 12% of the households in our sample, but does not vary by treatment status. Two thirds of this attrition is due to the security situation in some of the villages at midline, which were deemed too unsafe for the survey teams to enter. Attrition is independent of treatment status, and is not predicted by baseline characteristics.[26]

[23]As an example from the sub-domain of gross motor skills, the caretaker of a child aged 19-20 months is asked "Does the child run fairly well, stopping himself/herself without bumping into things or falling?". Each task is scored with: 0 points if the caretaker reports that the child does not perform it yet; 5 points if the child performs it "sometimes"; and 10 points if he/she does it habitually. The questionnaire was adapted to the Northern Nigerian context and translated into Hausa.

[24]As shown in Rubio-Codina et al. (2016), there are questions about the internal and external validity of some of the ASQ modules, especially for the fine motor domain.

[25]Given the likely measurement error in children's dates of birth, we adopt a loose definition for 'in utero', considering children who would have been aged -10 to 1 month at baseline.

[26]Table 2.A2 shows that attrition does not vary by treatment status, and is mostly attributable to insecure

2.3.4 Programme exposure

We present intent-to- treat estimates of the effects of the Child Development Grant Programme (see Section 2.4). To frame the results, we detail the exposure of the sample households to the two components of the intervention – cash and information.

Take-up of the cash component is high: Panel A of Table 2.3 shows that 83% of women in treated villages report ever having received the cash grant. Failure to demonstrate the pregnancy is the main reason provided by women for not participating – accounting for half of the remaining 17% – followed by programme implementation issues. Just 6% of women in control villages have accessed the grant. Some cross-village registrations have been documented in the process evaluation, whereby women pretend to be residents of treated communities to receive the grant (Sharp et al., 2016). Given that the midline data is collected around 2 years after baseline, most women are still receiving the cash transfer at the time of the midline survey. This introduces an element of censoring in the exposure.

We examine the timing of the cash grant with respect to women's pregnancies by merging the date of birth from our midline data with the administrative database used to schedule grant payments.[27] Around half of the sample actually starts receiving the cash grant while pregnant, with an additional 15% in the same month when the child is born (Panel B of Table 2.3). Still, around a third of women receive their first payment after the child is born. This can be explained by slower-than-expected programme rollout in some areas, and also by women who initially failed to enrol being included for a later pregnancy.

Finally, Panel C of Table 2.3 investigates the intensity of the cash transfer, using the information from the administrative database. At the time of the midline interview, women have received around 19 monthly transfers. As highlighted above, there is some censoring, as this number covers slightly more than 80% of the total number of payments that women can expect before their child turns two years old. The transfers cumulatively amount to an average of US$ 684 – with a median of 723 and a 90-10 range of 594-790. No differences emerge between T1 and T2 in the timing or intensity of the cash transfer.

We measure exposure to the information component by asking respondents to recall whether they have seen CDGP materials in their village, or attended any of the activities that are part of the Behaviour Change Communication strategy.[28] Almost all women in treated communities, and a vast majority of their husbands, recall being exposed to at least one of the low-intensity Behaviour Change Communication channels – as seen in Panel A of Table 2.4.

villages that could not be visited. Table 2.A3 shows that insecure villages are comparable in terms of availability of facilities and mean household wealth at baseline, but (expectedly) differ in the incidence of man-made shocks such as curfews or violence in the 12 months before baseline.

[27] Women are matched to the database using the number of the simcard they receive at enrolment. A second attempt at matching is made on the basis of names within each village. In the case of both matching approaches being unsuccessful, we use self-reported month of first payments.

[28] Measuring exposure to the information component has proved challenging. For some of the Behaviour Change Communication channels, it's not always possible to precisely elicit whether the respondent has indeed been exposed to CDGP activities or some other similar interaction. Respondents' recall might be incomplete, or could be hard to attribute to CDGP. As an example, posters affixed in health facilities and village centres, as well as radio ads and programmes, are not immediately identifiable as CDGP-related, especially by the less literate, and might be confused with materials from other public health interventions.

However, more than half the subjects in control areas also report being exposed. This is particularly true for channels that have limited excludability (such as posters or radio transmissions), indicating potential spillovers. On the other hand, health talks and food demonstrations are known to more than half of treated women, but 10% or less of control women (Table 2.A4).

High-intensity channels should only have been offered in T2 villages. In fact, Panel B of Table 2.4 shows similar levels of attendance across T1 and T2 for support groups and one-on-one counselling. Women in T2 villages are just 10-15pp more likely to have attended either activity. We suspect the main explanation for this finding is connected to the difficulty in accurately eliciting attendance to these activities. Programme monitoring data did not capture who had attended support groups and one-on-one counselling. We therefore needed to rely on self-reported attendance. Respondents might not be able to separately identify interactions with community volunteers that occur as part of the CDGP high-intensity information component, from other more informal contact with the Community Volunteers. We also observe very low participation of husbands: it has been suggested that the activities are viewed as mainly concerning women, and that men are less prone to engage with Community Volunteers on matters around raising children (Sharp et al., 2016).

2.4 Effects of CDGP on child development

The evidence above shows that – as expected – no differences emerge between the low-intensity (T1) and high-intensity (T2) treatment arms with respect to take-up of the cash component of the intervention. However, as far as we are able to measure, exposure to the information component was also very similar, with only minor differences between T1 and T2 even as far as the high-intensity channels are concerned. For the purpose of this paper, we present all results pooling T1 and T2 villages into a single treatment group, and comparing them to control (C) villages.[29]

We estimate the effects of the CDGP intervention on a range of household-, woman-, and child-level outcomes measured at midline, i.e. approximately 25 months after the baseline. We adopt an intent-to-treat (ITT) approach, exploiting the random assignment of villages to treatment arms. For outcomes that are observed at both baseline and midline, we adopt the following ANCOVA specification (McKenzie, 2012):

$$Y_{ivd,t_1} = \gamma T_{vd} + \boldsymbol{X}_{ivd,t_0}\beta + \delta Y_{ivd,t_0} + \eta_d + \lambda_s + \varepsilon_{ivd,t_1} \tag{2.4.1}$$

Where Y_{ivl,t_1} is the outcome at midline for household i residing in village v in district d; T_{vl} is the treatment indicator; $\boldsymbol{X}_{ivl,t_0}$ is a vector of controls – which include an intercept and a set of pre-treatment characteristics (family composition, total equivalised per capita household expenditure, and the woman's age, education, and marital status); η_d are district (LGA) dummies; λ_s are dummies for randomisation tranche; and ε_{ivl,t_1} is a random error term.

[29] Separate estimates by treatment arm do not reveal any relevant difference in the impacts of the programme. They are available on request from the authors.

We cluster all standard errors at the level of the unit of randomisation (village).[30]

Our choice of relying on an ITT approach is due to various factors. While receipt of the cash transfer can be accurately measured using a combination of administrative data and self reports, exposure to the information component is harder to assess at the individual level (see Section 2.3). In addition, we don't have at our disposal any other clearly defined source of exogenous variation besides the randomised assignment at the village level.

The presence of spillover effects between treatment and control areas might affect some of the results presented in this paper. While the cash has been disbursed in adherence to the randomisation protocol, some of the low-intensity BCC activities might have reached the control group – see Section 2.3. Moreover, even assuming perfect implementation of the BCC, evidence from the qualitative evaluation shows that information around infant practices tends to diffuse quickly across community boundaries (Sharp et al., 2016), indicating a 'social interaction' spillover (Angelucci and Maro, 2015). Moreover, we cannot exclude that contemporaneous campaigns by the government or other NGOs might have affected the control group beyond the CDGP. In light of the partial take-up of both cash and information, cross-village registrations, and likely information spillovers, our ITT estimates should be interpreted as lower bounds to the treatment effect on treated households.[31]

2.4.1 Fertility and gestation length

As a first step, we examine whether fertility has been affected by the CDGP intervention. Panel A of Table 2.5 shows that, among women who were pregnant at baseline, treated women had a slightly higher chance of having at least one live birth.[32] Around 13% of these women lost a live-born child in the period between baseline and midline, with no differences in infant mortality across treatment groups.

Mothers receiving the treatment have a longer gestation period. Children of mothers in the treatment group are born .55 calendar months (about 2 and a half weeks) later than control children. When controlling for possible differences between treatment and control groups in dates of interview at baseline and midline (Column 3), the difference persists. A similar difference of approximately half a month is observed when we focus on our main estimation

[30] For outcomes that were not observed at BL, such as the ones related to the new child, we drop the level Y_{ivd,t_0}:

$$Y_{ivd,t_1} = \gamma T_{vd} + \boldsymbol{X}_{ivd,t_0}\beta + \eta_d + \lambda_s + \varepsilon_{ivd,t_1} \tag{2.4.2}$$

[31] In the design of the intervention, no special adjustments were made to directly estimate the magnitude of cross-cluster spillover effects as suggested by recent literature (Baird et al., 2015). The lack of sources of exogenous variation prevents the use of quasi-experimental approaches relying on instrumental variables or discontinuities. One approach that could shed some light on spillovers is the estimation of treatment effect heterogeneity using villages that are more and less geographically isolated as control groups. Such investigation is beyond the scope of the current work.

[32] This finding might be explained by treated women intentionally seeking to get pregnant more often. However, our sample focuses on women who reported being pregnant before the programme started, so observed differences in the chance of giving birth should mostly be explained by subsequent pregnancies. Alternatively, the CDGP might have reduced the occurrence of miscarriages and stillbirths. We cannot directly confirm this or rule it out, as the midline questionnaire did not survey this aspect.

sample, i.e. the 'in-utero' sample, in Panel B of Table 2.5.

The gestation effect in Table 2.5 can be estimated thanks to the unique design of our intervention, starting in pregnancy. This is a novel finding, which to our knowledge has not been noticed or investigated in similar studies or settings. While there is an abundance of studies documenting which factors influence birth weight, the evidence on gestation length and preterm birth is more scant. In particular, there is very little literature looking at determinants of gestation length in developing countries. This might be because gestational age is hard to estimate in a developing country context. Ultrasound measurements (the gold standard for the antenatal period) require expensive equipment, while assessments at the time of birth based on clinician-administered scores – such as the Ballard and Dubowitz scores – require technical skills that might be absent in low-income settings. Furthermore, these scores might not be valid in populations characterised by malnourishment, due to different patterns of intrauterine growth (Rosenberg et al., 2009).

Unfortunately, our data does not allow to construct a direct estimate of gestational age. We thus cannot establish which children in our sample are born preterm – i.e. at less than 37 weeks into the pregnancy. Furthermore, birth weight is not available in our data either: children in the areas we survey are rarely weighed at birth (they are often born at home), and birth registries have little to no coverage. Nevertheless, some estimates from Subsaharan Africa are available to serve as a benchmark for our finding of a two-week gestation effect. Two hospital-based studies for Nigeria (Omigbodun and Adewuyi (1997) in Ibadan, and Okeke et al. (2014) in Enugu) estimate gestational age at birth to be 280 and 275 days respectively. Estimates of the standard deviation of gestational age in Omigbodun and Adewuyi (1997) vary by age and parity, but are all around 15 days. This is consistent with evidence from Malawi (Verhoeff et al., 1997) and Sudan (Elshibly and Schmalisch, 2008). This would put our estimated size of the gestation effect at 1 SD of comparable populations.

Gestation length and preterm birth have been linked to to various maternal factors, such as age, nutritional status, infection, mental health, and lifestyle (Althabe et al., 2012). Smoking and drinking during pregnancy have long been identified as a hazard for birth outcomes, including gestation length. Rosenzweig and Schultz (1982) notice a negative association between smoking and gestation length of slightly less than half a week. Wen et al. (1990) estimate this association as 0.5-1 week for mothers over 30. Li and Poirier (2003) set up a simultaneous equation model for NLSY data to account for endogeneity of maternal inputs, and find an insignificant effect of smoking on gestation length, but drinking reduces gestational age at birth by 1.2 weeks. Estimates of the effect of prenatal care on gestation length vary widely, from null (Evans and Lien, 2005) to more than 2 weeks Li and Poirier (2003). A recent review concludes that the evidence of the effects of prenatal care on birth outcomes is still far from conclusive (Corman et al., 2018). Finally, pollution seems to be causally related to gestation length (see Currie et al., 2009 and references therein).

Interventions and policies in developed countries have also shown mixed success at improving gestation length and reducing the incidence of preterm births. Early evidence from the WIC programme in the US, which provides advice and nutritional supplementation to low-income pregnant women, shows positive effects on birth weight. It also points to increased gestation length, with associations ranging between .25 and .75 weeks (Devaney et al., 1992; Bitler and Currie, 2005); however, more recent re-assessments (Joyce et al., 2005, 2008; Hoynes et al., 2011) show that the association between WIC and gestation is mostly due to spurious associations resulting from programme enrolment, and becomes negligible once selection bias is accounted for. These results are in agreement with recent research on the effect of cash or in-kind transfer programmes on birth outcomes.[33] Amarante et al. (2016), after showing similar estimates for a cash transfer in Uruguay, argue that in most cases transfer programmes increase birth weight predominantly by improving intrauterine growth, rather than by lengthening gestation. However, the evidence is not incontrovertible: Hoynes et al. (2015) show that the Earned Income Tax Credit in the US does reduce the incidence of preterm birth.

2.4.2 Physical growth

CDGP led to improvements in height for children born into the intervention. At the same time, it resulted in increased wasting (low weight-for-height). We focus here on the new children – i.e. those who were 'in utero' at baseline, and are aged 14-27 months at midline. We derive age-normed scores for height and weight based on a healthy reference population, using standard growth charts (WHO, 2009).[34] Intention-to-treat results are in Table 2.6. The intervention has increased child height-for-age by .21 standard deviations of the reference population, while having no effect on weight or arm circumference. To put the height effect into perspective, the difference in mean height for age between the top and bottom wealth quartile – as measured by the PPI index (Schreiner, 2015) – is around .25 in the control group. Treated children, being taller but not heavier, have a reduced weight-for-height. Correspondingly, there is also a reduction in stunting (low height-for-age) of about 4.8 percentage points, but an increase in wasting (low weight-for-height).

How can this seemingly counterintuitive result be explained? Stunting usually reflects long term, chronic undernutrition. Wasting is instead viewed as a symptom of acute undernutrition.[35] CDGP might be effective in relieving some of the early causes of chronic malnutrition throughout the first thousand days, such as suboptimal breastfeeding and complementary feeding practices, thus promoting height. At the same time, the intervention might not be able to tackle other systemic causes of malnutrition – e.g. lack of secure food and water supply, or poor hygiene, sanitation, and health services. Thus, weight gains might not keep up with height gains. Another explanation is connected with the timing of the fieldwork. Households are visited during the rainy season, around the time of harvest. This is immediately after the leanest time of the year, where food is scarce and prices are high. If wasting approximately reflects undernutrition in the previous six months, the effect of the lean season might compress

[33] For example, improvements in birth weight have been found for the *Oportunidades* conditional cash transfer programme in Mexico (Barber and Gertler, 2008) and the Food Stamps Program in the US (Almond et al., 2011).

[34] We compute z-scores using the Stata routine `zscore06` (Leroy, 2011).

[35] While it's true that wasting has a more transient nature, wasting and stunting are likely to be observable manifestation of similar physiological processes. In particular, decrease of muscle and fat mass is thought to underlie both stunting and wasting (Briend et al., 2015).

the distribution of child weight to the bottom while preserving the height gain.

It's important to interpret the physical growth results in light of the gestation effect in Table 2.5. This causes treated children to be younger than control children when measured at midline. Height-for-age (HAZ) scores in deprived populations exhibit a downward slope during the first 24 months of life – a phenomenon known as growth faltering. We observe growth faltering in our sample; the top panel of Figure 2.1 shows that mean HAZ decreases approximately from -2 to -3 in the 14-27 months age range for our sample of new children, while weight-for-age (WAZ) has a slightly more stable profile. This is consistent with Shrimpton-Victora curves for Subsaharan Africa (Shrimpton et al., 2001; Victora et al., 2010).[36] Thus, our estimated HAZ effect of .21 is to be interpreted as the sum of two components: an *at-age* effect, whereby treated children are taller than control children at the same age, and a *composition* effect, whereby treated children are younger and thus slightly to the left of control children on the growth faltering profile.

While the raw effect is interesting per se, we seek to shed more light on the relative importance of the composition effect. We find that the actual at-age effect is positive, and dominates the composition effect. The last column of Table 2.6 shows effects on z-scores after controlling for child age. This excludes composition effects and isolates the at-age component. Indeed, the estimate on HAZ is lower than its raw counterpart at .12, but still well above zero.[37]

As an additional check, we implement an instrumental variables strategy that treats age as endogenous. We account for the potential endogeneity of age by instrumenting it with exogenous variation in interview date. The midline data were collected over a period of two months, from early October to early December 2016. Children in households that were surveyed later in this period will be comparatively older than children surveyed earlier. The instrument is relevant, as assessed in the first-stage regressions displayed in Table 2.A7. We use calendar day of midline interview for our main specifications. The validity of the instrument relies on the assumption that fieldwork decisions around when to visit each village are orthogonal to unobserved determinants of child growth.

Control function estimates using date of interview are presented in Table 2.A8. There is no difference in the estimates for weight and arm circumference. For height, control function estimates appear to be slightly higher but very similar to the age-adjusted effect, independently of the discretisation we adopt for the age control (see Table 2.A11). This is consistent with a small downward bias of the OLS estimate due to endogeneity. Again, we interpret the similar size of the control function effect to the age-adjusted effect as an indication that most of the raw height effect is driven by the 'at-age' component rather than by the small difference in age between treated and control children. In Figure 2.1, control function estimates by age group are depicted jointly with the raw estimate.

As a final robustness check, we analyse height impacts using unstandardised and internally standardised measures of height, as seen in Table 2.A9. Our internal Z-score is obtained

[36] Even more similar to the curves in Victora et al. (2010) are the age profiles of z-scores for the old children surveyed at baseline, shown in Figure 2.A4.

[37] This type of estimate is robust to different polynomial and discrete adjustments for age, as seen in Table 2.A10.

by standardising height in centimeters at each month of age using mean and variance profiles smoothed by kernel-weighted local means. The results are consistent with the above interpretation. To see why, consider that: HAZ (computed according to the WHO growth standards) exhibits a downward slope in age because of growth faltering; the treated children, being younger, will look taller. Height in cm is naturally increasing in age; the treated children will look shorter. Our internal Z-score is by construction not a function of age. Thus, adjusting for age will dampen the effect on HAZ, increase the effect on height in cm, and will not influence the internally standardised score. This is exactly what is observed in Table 2.A9.

We also observe anthropometrics for children who were born before the programme had been rolled out. These children were exposed to the CDGP intervention only after birth. The improvements in diet diversity (see Table 2.A28 and Table 2.A27) and health (see Table 2.A25 and Table 2.A24) are very similar to the ones of their younger siblings. However, there is no effect on the anthropometrics of these older siblings (Table 2.A12).[38] This result directly speaks to the recent literature on growth faltering and catch-up growth (Anand et al., 2018). It seems to underscore that intervening early in the first thousand days – including the pregnancy period – might be necessary to obtain growth improvements. Despite similar improvements in diet, these children were too old to benefit from the intervention during their first months of life, when breastfeeding practices and the appropriate introduction of solid foods play a role. Curiously, this result is somewhat opposite to Levere et al. (2016), where positive weight effects occurred for older siblings of the children targeted by the intervention.

We place our findings on child growth in the context of the existing literature in Figure 2.5. We compare the findings in this paper with ones from other randomised evaluations taking place in low- and middle-income countries. The interventions listed in the figure are very heterogeneous in terms of socioeconomic and cultural context, timing, scope, target population, and components. Comparisons are thus to be entertained with caution. Still, the size of our height effect is comparable to the largest non-null effects observed for conditional cash transfers. The null effect on weight appears to be a common finding in the literature.

2.4.3 Communication and motor skills

The CDGP had a small positive effect on communication skills but no effect on gross motor skills. This was measured in our survey using the Ages and Stages Questionnaire (ASQ). In Table 2.7 we present effects on standardised scores (using the mean and standard deviation from the control group) and on binary indicators of children's scores falling in the normal range, i.e. above the thresholds for development problems. In the absence of locally validated thresholds, we use the ones from the reference western population. Like in the case of anthropometrics, we adjust the estimates for the small age difference between treated and control children. We also flexibly adjust the estimates to account for the specific age bands at which different modules of the ASQ questionnaire are asked. The CDGP intervention had

[38]We present these results in terms of unstandardised height, weight, and MUAC, as the standard growth charts run up to 59 months, while 40% of the old children are older than 59 months at midline.

a marginally significant effect on the communications skills score, of around .1 of a standard deviation. It also increased the proportion of children with normal communication scores by more than 5 percentage points.

To the best of our knowledge, our study is the first to investigate the impacts of a large-scale cash transfer on child cognitive and motor development in Sub-saharan Africa. In general, only a few randomised evaluations in developing countries have surveyed these dimensions.[39] Given the wide range of instruments and scales used in these studies, and the different domains of child development they cover, our results are only partially comparable to the existing evidence. Our null finding on gross motor skills is the consistent with Macours et al. (2012) and Levere et al. (2016), while our small effect on communication skills is similar in size to what is found for other cognitive domains such as vocabulary and memory.

2.4.4 Child health

The health status of children has also improved as a consequence of CDGP (Table 2.8). Children born after midline are more likely to receive vaccinations, in particular BCG, measles, and yellow fever.[40] They are also significantly more likely to have received deworming medication, less likely to have been ill, and less likely to have suffered from diarrhoea. Similar effects are observed on their older siblings. Still, the proportion of children affected by illness and diarrhoea remains staggeringly high.

2.5 Mechanisms

In this section, we examine potential channels that may explain the observed effects of CDGP on child development and health. The depth of the survey allows us to investigate a range of potential mechanisms, encompassing knowledge and practices around breastfeeding, child diet, and household resources.

2.5.1 Infant and young child feeding knowledge and practices

Health and nutrition practices in the first thousand days of life have the potential to significantly impact physical and mental development. In the 0-6 months range, breastfeeding has a central role; prompt initiation of breastfeeding reduces infant mortality (Debes et al., 2013), and

[39] Paxson and Schady (2010) measured children's vocabulary, memory, visual integration, and socioemotional skills. Macours et al. (2012) fielded a number of tests from different batteries covering memory and motor skills, together with a vocabulary test and the Behaviour Problem Index to measure socioemotional/behavioural skills. Levere et al. (2016) uses the same ASQ instrument as we do, but for a wider range of domains including fine motor and social-personal skills.

[40] We collect data on vaccinations by first asking for any vaccination card to be shown. Only around 12% of mother in our main sample can show a vaccination card for their new child; in case the card cannot be produced, we elicit maternal reports. Details of programme effects on vaccinations are in Table 2.A14. Coverage of polio immunisations was unaffected, as it is already high in the control group, but large improvements can be observed in most other types of vaccination. Nevertheless, a negligible proportion of children have received all recommended vaccinations.

exclusive breastfeeding can prevent child morbidity (Kramer and Kakuma, 2012). Informational and educational interventions have proved effective at promoting correct breastfeeding practices (Imdad et al., 2011; Haroon et al., 2013).

Our findings show large impacts of the programme on knowledge, attitudes, and practices regarding infant and child feeding, with especially large increases reported in exclusive breastfeeding rates. In accordance with global guidelines (WHO, 2008), the CDGP curriculum recommends initiating breastfeeding immediately and breastfeeding exclusively until the child turns 6 months old – see Table 2.1. We broadly categorise early childhood practices into pre-, peri-, and postnatal. Estimates of the effects on parental knowledge around infant and young child feeding – in Table 2.9 – are large and positive for all indicators, and for both the mother and her husband.[41] Surprisingly, baseline knowledge of husbands seems to be slightly better than their wives' for some indicators. The rate of "correct" responses on questions related to early and exclusive breastfeeding (Behaviour Change Communication messages 3 and 4 in Table 2.1) increases markedly from very low baseline levels. Table 2.10 presents programme effects on actual child practices, as reported by the mother with reference to the new child (the child born just after the baseline). Similarly to the knowledge results above, large impacts are observed for practices related to timely breastfeeding and antenatal care.[42]

These indicators of knowledge and practices are derived from direct questions to respondents; they should thus be interpreted with caution, due to potential self-reporting bias. However, qualitative evidence from unstructured interviews with a subset of beneficiary households indicate widespread understanding of the practices recommended as part of the Behaviour Change Communication component; respondents report embracing the suggestions after observing their beneficial effects on children in their community (Sharp et al., 2018). Furthermore, in the likely presence of spillover effects due to social interactions across communities, these ITT estimates constitute a lower bound for the true treatment effects on treated households.

Complementary feeding refers to the introduction of safe and nutritious solid and semi-solid foods alongside breast milk, in the 6-24 month age range. The quality and diversity of complementary feeding is linked to improved growth and morbidity (Lassi et al., 2013). In particular, animal source foods seem to play a very important role in enabling children to achieve their physical growth potential (Headey et al., 2017). In our survey, a 24-hour food recall module was used to measure dietary diversity for the new child and (when present) the

[41] WHO guidelines suggest to breastfeed exclusively for 6 months. The indicator "Best to breastfeed exclusively for 6-7 months" is constructed to allow for answers reported in weeks to still be considered appropriate if they slightly exceed 6 months.

[42] Increase in antenatal care seeking behaviours are not confined to prevalence. Treated women also receive .63 more antenatal care visits, as seen in Panel A of Table 2.A15. Panel B, focusing on women who were pregnant when surveyed again at midline, shows that this increase continues into later pregnancies: treated women are more likely to have already received antenatal care and to have received iron supplements. There is no evidence that this is driven by supply-side improvements in antenatal care, since the cost of treatment and transport is unchanged. More evidence about the absence of healthcare supply-side effects is in Table 2.A16: the availability of services and staff at the nearest health facility to each village is unaffected by treatment status – apart from healthy diet counselling which is probably picking up the presence of community volunteers implementing the educational component of CDGP in treated villages.

old child. The caregiver is asked to list every item of food the child consumed in the whole day previous to the interview. Then, for each food item, she is asked to provide all ingredients used to prepare it. These are then categorised into food groups. No attempt was made to elicit the quantity of food fed to the children. We construct a diet diversity index based on the number of food groups following the approach in WHO (2008). Estimates of the effect of CDGP on child diet are in Table 2.11.

New children in the treatment group consumed on average .35 more food groups in the day prior to the midline interview than control children. Overall consumption of fruit and vegetables was unchanged, but a shift is visible from dark leaves (spinach, kale, chard) to vitamin A-rich fruits and vegetables (e.g. mango, pawpaw, pumpkin, sweet potato). This result does not have a straightforward positive interpretation, as dark leaves are found to carry important nutritional value in the complementary feeding phase. Potential explanations for this shift might be connected to local beliefs about the relative nutritional value of these two food groups: it might be that dark leaves are perceived to be inferior, lower quality foods that are increasingly abandoned as incomes rise in favour of orange-flesh fruit and vegetables.

Finally, there was a 15pp increase in children consuming animal-source foods, driven especially by dairy products. Effects are similar for the older siblings, although the increase in animal protein consumption is more sizeable for the new children. This is an important result given the recent evidence on the role of animal protein in the growth process.

2.5.2 Resources and productive impacts

Besides supporting households' livelihoods in the short term, in some instances large and long-lasting cash transfers have productive effects on expenditure, labour supply, and investment patterns. These effects can multiply the benefits of the transfer, and extend the time horizon in which their economic benefits are realised, leading to long-term improvements in living standards. Recent long-run evaluations of interventions in similar contexts have shown promising results – see for example Gertler et al. (2012) for Mexico, Handa et al. (2018) for Zambia, and Daidone et al. (2017) for an overview of seven different interventions in Subsaharan Africa. Notably, productive effects are not necessarily restricted to receipt of cash: the information-only intervention evaluated in Fitzsimons et al. (2016) caused an increase in male labour supply.

Our data from the two-year midline followup does not allow us to investigate long-run effects. However, even at this early stage, the CDGP intervention had noticeable effects on household resources and livelihoods. Table 2.12 displays effects on expenditures, borrowing, saving, and food security.

The transfer leads to substantial increases in monthly household expenditure, with most of the transfer being spent on food. Total monthly household expenditure increased by an amount comparable, if not superior, to the grant – whose PPP value was 21.6 USD. Most of the increase in absolute terms, and the entire increase in equivalised terms (Panel B), is attributable to food expenditure. This is consistent with respondents' reports on how the grant money was used: almost 90% of women and husbands say that most of the cash they received from CDGP was spent on food for the children, or for the household in general – see Table 2.A19. Beneficiaries don't seem to be changing their savings or borrowing in response to receipt of the grant, as seen in Panel C of Table 2.12.[43] Again, this is consistent with households spending the entire grant amount, without saving any at least on average.

CDGP increased expenditure on animal-source foods, again particularly on meat and eggs. As seen in the previous subsection, children in treated households have a more varied diet, with increased consumption of animal-source foods. The nutrition results are mirrored in food expenditure patterns at the household level in Table 2.13.[44] Furthermore, CDGP also improved food security throughout the year. Food security had been worsening in Northern Nigeria in period leading up to the CDGP midline, due to factors such as depreciation of the national currency and violent conflicts (FEWS, 2016). Households reporting not having enough food to eat in the 12 months previous to the interview were 14% of our main sample at baseline in 2014, with no difference between treated and controls. Two years later at midline, almost 30% of control households had gone without enough food in the previous 12 months. The effect of CDGP on food security is thus to be interpreted as partially offsetting this worsening outlook.

The intervention did not affect husbands' work activities on any margin that we can observe in our data. Table 2.14 shows the effect of CDGP on the work activities of women and their husbands. As the main providers of food for their household in a society characterised by subsistence farming, almost all husbands of the women in our sample are employed in agriculture. On average, they also have a second work activity.

On the other hand, work participation of women who were pregnant at baseline has increased by more than 7 percentage points, mostly driven by increased uptake of 'self-employed' activities. Our data does not allow us to investigate which specific kind of activity has been incentivised by the programme. As highlighted in a previous section, home segregation during daytime limits which kind of activities women can undertake. For example, they are normally not allowed to farm open land. However, evidence from the qualitative workstream that is taking place as part of the evaluation shows that women were able to invest into small-scale home-based activities such as petty trade, food processing and sale, small livestock rearing, and services to other women (e.g. hairdressing or pounding grain) (Sharp et al., 2018).

Effects are also observed on the intensive margin. Treated women engage on average in .11 more different work activities, and there is a small effect on the number of days usually worked at their highest-earning activity. Despite the inherent challenge of measuring earnings in this kind of context, we detect a small and marginally significant effect on reported monthly earnings from work activities excluding crop sales.

[43] Additional results on saving and borrowing are in Table 2.A17.

[44] No effects are observed on home production of food, as seen in Table 2.A18. We thus exclude that dietary diversity improvements arise because of different household choices in home production.

2.5.3 The role of information

The design of the evaluation does not allow us to directly disentangle the effect of each component. Nevertheless, in this section we provide some evidence to argue that the observed effects on child development are not entirely explained by the provision of cash, and that the information component had at least some separate effect on household behaviour. We do this by: (i) undertaking a mediation analysis to decompose the effects on child development; and (ii) examining the programme's impact on Engel curves for food expenditures.

To further explore the relative role of information, practices, and resources in explaining the effects on child growth and development, we use mediation analysis. We follow the approach in Gelbach (2016), which allows us to decompose ITT coefficients into components explained by different groups of potential mediators in a way that is invariant to the order in which these mediators are considered. The decomposition is based on the well-known omitted variables formula: a series of linear models are estimated, with and without the inclusion of each group of mediators.[45] We cannot assign any causal interpretation to our mediation results. On its face, the analysis might resemble an exercise in production function estimation (Cunha et al., 2010; Attanasio et al., 2017). However, we do not have at our disposal enough sources of plausibly exogenous variation that would allow us to identify and recover any technology parameters. Our mediation analysis is simply seeking to characterise which of the potential channel we observe is contributing more significantly to the overall effects on child development.

We find that maternal knowledge, antenatal care practices and dietary diversity were key mechanisms driving the effects we observed on child height and communication skills. Figure 2.6 depicts the relative contribution of different groups of mediators. We focus on the age-adjusted effects on height-for-age and communication skills, since we observe no effect for gross motor skills. In the case of HAZ, the mediators explain more than 60% of the ITT effect. Approximately half of the effect can be attributed to the maternal knowledge index in Table 2.9. The only other significant mediator group is antenatal care practices. For communication skills, each of knowledge, antenatal care practices, and dietary diversity explain 20% of the ITT – although only the latter two are significant.

For both HAZ and communication skills, purely resource-based mediators – household expenditure and maternal labour supply – explain a negligible fraction of the ITT. Most of the ITT goes through knowledge, or practices like breastfeeding, antenatal care, and dietary diversity. These dimensions are either independent of resources, or could presumably be affected by both resources and information. As an example, correct breastfeeding practices can be though of as cost-free.[46] Dietary diversity, on the other hand, has a monetary cost. But it will only increase if the household has more resources to spend on food *and* its members

[45] More details on the decomposition are in Appendix Section 2.9.2. The full results underlying Figure 2.6 are in Table 2.A20.

[46] A possible alternative explanation is that better nutrition ensured by the cash component enables malnourished mothers to breastfeed more effectively. We can't explore this aspect in our data, but it never emerges as a significant factor in the qualitative evaluation (Sharp et al., 2018).

value diversity in their children's diet. This underscores how it is likely that the intervention's information component indeed contributed to improving child development.

2.5.4 Food expenditure patterns

As a final piece of evidence, we turn to estimation of Engel curves. These curves describe the relationship between the share of household expenditure devoted to food and household expenditure itself. Omitting village and region indices for simplicity, a simple linear form for household *i*'s food Engel curve is:

$$W_i = \alpha_1 E_i + \alpha_2 T_i + \alpha_3 E_i * T_i + \boldsymbol{X}_i \beta + u \tag{2.5.1}$$

where T_i is the treatment indicator, E_i is total nondurable expenditure, W_i is the share of food expenditure over total nondurable expenditure, and $\boldsymbol{X}_i$ is a set of covariates including an intercept. The parameter α_1 establishes how food share varies with total expenditure for the average household. If food is a necessity good, α_1 is negative. The parameters α_2 and α_3 allow for the intervention to change the shape of the Engel curve, by shifting it intercept and its slope, respectively.

Total nondurable expenditure is very likely to be endogenous in (2.5.1), due to measurement error and possibly to correlation with unobserved preferences that might determine food expenditure. A common solution to endogeneity in this literature is to instrument total expenditure using information about the household's wealth (Armand et al., 2016). In our case, we use the PPI wealth index, a summary measure of wealth based on family composition, assets, and dwelling features (Schreiner, 2015).

Results are in Table 2.15. Column (2) shows ITT effects of the intervention, while columns (3) and (4) show estimates for the parameters in Equation (2.5.1). The effect if CDGP on overall food share is positive and around two percentage points, albeit insignificant. Food is a necessity, with a decreasing share conditional on treatment status. When the slope of the Engel curve is restricted to be independent of treatment, the CDGP increases the share spent on food by 3.5pp at any given level of expenditure. Standard errors in the specification allowing for slope change are much larger, but estimates are broadly similar. These estimates describe a twofold effect of CDGP on food expenditure patterns: on one hand, the intervention – presumably through its cash component – has shifted household expenditure *along* the Engel curve; on the other hand, it has shifted the Engel curve itself upwards. The latter shift offsets the former, yielding a small insignificant programme-induced increase in the overall share. For animal-source foods, we observe a significant effect on the share of almost 4 percentage points. However, the Engel curves show how this increase mostly derives from meat, fish, and eggs being “luxury" foods, for which the Engel curve is upward-sloping.

Our findings are consistent with the literature. Many cash transfer programmes have highlighted this somewhat counterintuitive phenomenon: large increases in resources seem to not affect the share of household budget devoted to food, or in some cases even increase it. This runs counter to evidence from developed countries, where food share declines with household expenditure. In these cases, the income effect of the cash transfer seems to be offset by other features of the programmes in question that cause a shift of the Engel curves (Schady and Rosero, 2008; Attanasio and Lechene, 2010; Attanasio et al., 2012).

A common explanation for this phenomenon is the empowerment of women in recipient households. In a unitary model, where the household acts as single decision maker, it's irrelevant who the recipient of the transfer is. Women – who often receive the cash transfer directly – might have different preferences over expenditure. As long as they are able to at least partially retain control over the cash transfer, the relative rebalancing of household preferences can offset the income effect. In fact, recent evidence rejects the unitary model in favour of a collective alternative (Attanasio and Lechene, 2014; Armand et al., 2016). In the case of CDGP, we directly elicited measures of intra-household decision making power before the intervention. We asked the same set of questions about intra-household decision making to both the woman and her husband, at baseline. Figure 2.8 shows that men have predominant control about major household purchases and what food to grow and buy. More than half of women say they would be able to decide how to spend income and gifts accruing to them directly. Husbands' reports are very similar. When interviewed at midline, the vast majority of women and their husbands say that the woman is in control of how the cash grant is spent. This underscores how the reallocation of decision making around food expenditure towards women can at least partially explain the upward shift of the Engel curves.

However, the offsetting of the income effect could also be explained by informal or indirect conditionalities, often referred to as 'soft conditionalities'. Even when cash transfers are unconditional, recipients might be exposed to messaging around the 'best use' for the additional income; alternatively, the money received could be implicitly labelled for certain uses (Pace and Pellerano, 2016).[47] Both dynamics are involved in our intervention. There is implicit labelling of the grant towards the new child, and explicit messaging in the Behaviour Change Communication component.

2.6 Conclusion

Many children in developing countries fail to reach their development potential. Suboptimal development has long reaching consequences, stretching into the school age and adulthood. Intervening early in life might partially remedy this disadvantage. However, it's still not clear what type of intervention (or combination of interventions) is the most effective at tackling underdevelopment.

In this paper, we assess the impact of a 'cash plus' intervention in Nigeria on child development, exploiting an experimental evaluation design. Our intervention combines a cash transfer with a behaviour change educational component, and targets mothers and their children starting from pregnancy. We find that the programme made children taller for their age, but their

[47] Some examples are *Familias en Accion* in Colombia, the Child Grant Program in Lesotho (Pace and Pellerano, 2016), or Tayssir in Morocco (Benhassine et al., 2015). Behavioural labelling effects have also been documented for rich countries – see for example Beatty et al. (2014).

weight was unchanged – implying they are thinner. The estimated height effect covers 4 to 8 per cent of the gap between the children in our sample and a healthy reference population. Our intent-to-treat approach, while straightforward, returns a lower bound for actual treatment effects on the treated. A back-of-the-envelope inflation by the proportion of women in the sample receiving the cash grant (83%) would put the fraction of height-for-age gap covered at 5 to 10 percent. Treated children also have better communication skills, while no effect is observed on their motor development. No improvements are observed for their older siblings, exposed to the intervention at an older age – indicating a potentially important role for early intervention to remediate stunting.

The main limitation of our study is that the experimental design does not allow to neatly separate the effect of the cash and information components. However, we provide substantial indirect evidence that information played a separate role from cash. The CDGP led to across-the-board improvements in intermediate outcomes that do not hinge on household resources, such as knowledge and practices around infant and young child feeding. It increased dietary diversity, particularly in animal-source foods. We show that these changes in knowledge, practices, and diet explain a large part of the effects on height and communication skills. Finally, we show that the programme induced changes in food expenditure that seem to go over and beyond what a plain injection of purchasing power would imply. If such changes in behaviour are internalised by respondents, they might prove long lasting and capable of sustaining improvement even when women graduate out of the cash transfer on their child's second birthday. The four-year endline followup, planned for collection in late 2018, will further inform about long run dynamics.

Viewed in the context of the literature, our work has important implications for the design of nutrition-sensitive interventions. Cash transfers have proved effective at improving various outcomes from education to poverty and employment. However, their record on fostering child nutritional status is mixed. Among recent systematic reviews, Manley et al. (2013) show no overall effect of cash transfers on height, and Bastagli et al. (2016) find that only 4 out of 10 studies estimating effects of cash transfers on anthropometric indicators find any statistically significant improvement. Why is this? To put it simply, cash alone might not be enough to make a dent in the issue of child malnutrition. More generally, the causes of malnutrition are likely to be complex and intertwined, and addressing just one might not make any appreciable difference. For example, a community based health care and supplementation intervention in the same Northern Nigerian setting, failed to achieve any gain in child growth (Hansford et al., 2017). Programmes to improve water, sanitation, and hygiene rarely affect child nutritional status (Headey and Palloni, 2018). An ideal intervention would integrate cash with other similar components, such as psychosocial stimulation, counseling on behaviour change (breastfeeding, complementary feeding, hygiene practices) and supply side investments (clean water, improved health care and immunisation services).

In this perspective, the Child Development Grant Programme is a step in the direction of 'cash plus' integration. This might explain why our height effect is not trivial, and is on the high end of what is estimated for similar randomised evaluations (see Figure 2.5). It is also consistent with previous literature finding larger marginal impacts of cash transfers in more deprived areas (Manley et al., 2013). The region where our intervention takes place is arguably among the poorest settings investigated in comparable studies, with staggering poverty, endemic food insecurity, and low coverage and quality of healthcare. Moreover, our results show that nutritional improvements can be achieved with unconditional cash. If cash is accompanied by implicit labelling and Behaviour Change Communication messaging, resorting to expensive conditionalities might not be needed.

The complex, multicausal nature of human capital development in early life is increasingly recognised. In the terminology of production functions, the process is characterised by significant self- and cross-productivities between investments and human capital (Cunha and Heckman, 2007). Deepening our understanding of the development process and of the complementarity between investments is an important priority for further research. This can allow more effective design of interventions that can help address the root causes of deprivation in the developing world.

2.7 Tables

Table 2.1: Information Components of the CDGP

PANEL A: Key Messages

Period		Message	Details
Prenatal	1	Attend antenatal care	Attend antenatal care at least four times during pregnancy.
	2	Eat one additional meal during pregnancy	Eat one extra small meal or snack each day to provide energy and nutrients for you and your growing baby.
Perinatal	3	Breastfeed immediately	Start breast feeding your baby within the first 30 minutes of delivery. Colostrum is good for the baby.
	4	Breastfeed exclusively	Breastfeed your child exclusively until six months old. Do not give water, tinned milk, or any other food.
Postnatal	5	Complimentary feeding	Introduce complimentary foods at six months of age while continuing to breastfeed. Breastfeed on demand and continue until two years of age. Gradually increase food variety as the child gets older.
	6	Hygiene and sanitation	Wash your hands after going to the toilet, cleaning baby who defecated, before and after feeding baby; wash baby's hands and face before feeding.
	7	Use health facilities	Take baby to health facility if you notice any of the following: fever, convulsion, refusing to eat, malnutrition, diarrhoea.
	8	Nutritious food	Ensure you buy nutritious foods when you are buying food for your family.

PANEL B: Channels

Type	Channel	Details
Low-Intensity	Information and education posters	Health and nutrition related posters are affixed in health facilities and village centres.
	Radio jingles / phone-in programmes	Jingles are played regularly on local radio channels. Phone-in programmes are one-hour shows in which CDGP staff and invited experts talk about one selected topic, and listeners can call in with questions.
	Friday preaching / islamic school teachers	
	Health talks	Trained health workers come to the village and deliver a session on a selected topic, with the aid of information cards. Any village resident can attend these talks, irrespective of beneficiary status.
	Food demonstrations	CDGP trained staff delivers nutrition education about the benefits of different foods, and demonstrates how to prepare and cook nutritious meals for children and other household members.
	Voice messages	Pre-recorded messages are sent to beneficiaries' programme phones to reinforce key messages.
High-Intensity	IYCF support groups	Groups are formed within communities to support beneficiaries, under the supervision and facilitation of community volunteers and health extension workers. The recommended size is 12-15 people, meeting once a month. They are also offered to men.
	One-on-one counselling	Beneficiaries and their husbands can consult community volunteers on an 'as needed' basis to receive specific information and training.

Notes: Panel A lists the eight key messages around which the behaviour change communication component of CDGP was built. Panel B details the channels by which the key messages were delivered to beneficiaries in treated villages.

Table 2.2: Baseline Balance

Sample: In Utero Pregnant Women at Baseline
Means, Standard Deviation in Parentheses, p-values in Brackets

	(1)	(2)	(3)	(4)	(5)	(6)
	Control	**T1**	**T2**	**pvalues of difference**		
	Mean (SD)	Mean (SD)	Mean (SD)	T1-C	T2-C	T2-T1
Household						
Observations	712	743	761			
Household size	7.63 (4.24)	7.38 (4.20)	7.55 (4.30)	[.378]	[.898]	[.447]
Progress out of Poverty Index (PPI)	26.52 (12.58)	28.87 (14.75)	26.99 (12.94)	[.064]	[.730]	[.108]
Equivalised daily per capita exp. (USD PPP)	1.52 (1.74)	1.84 (2.47)	1.59 (2.06)	[.091]	[.788]	[.134]
Woman and husband total monthly earnings (USD PPP)	160.29 (323.80)	185.12 (339.50)	166.13 (349.20)	[.259]	[.820]	[.405]
Did not have enough food at some point in past year	0.138	0.137	0.155	[.996]	[.533]	[.508]
Woman						
Observations	712	743	761			
Ever attended school	0.177	0.223	0.209	[.201]	[.379]	[.654]
Age (years)	25.70 (6.68)	25.24 (6.44)	25.53 (6.72)	[.241]	[.692]	[.451]
Num. children aged 0-2	0.47 (0.51)	0.47 (0.50)	0.48 (0.53)	[.953]	[.783]	[.712]
Months pregnant at baseline	5.48 (2.18)	5.46 (2.07)	5.29 (2.15)	[.816]	[.108]	[.170]
In polygamous marriage	0.490	0.440	0.515	[.119]	[.358]	[.014]
Any Work activity	0.756	0.693	0.710	[.142]	[.373]	[.583]
Has a say on major HH purchases	0.480	0.510	0.480	[.427]	[.926]	[.476]
Has a say on what food to buy	0.428	0.447	0.402	[.619]	[.495]	[.220]
Husband						
Observations	710	737	758			
Can read and write	0.652	0.613	0.649	[.390]	[.901]	[.432]
Ever attended school	0.555	0.537	0.499	[.626]	[.122]	[.318]
Age (years)	39.38 (10.85)	39.21 (12.07)	39.36 (10.35)	[.824]	[.994]	[.813]
Cultivated land	0.962	0.945	0.954	[.328]	[.610]	[.568]
Any Work activity	0.947	0.946	0.929	[.792]	[.433]	[.343]
Old child						
Observations	528	547	575			
Age (months)	36.73 (11.29)	37.13 (11.61)	36.20 (11.54)	[.498]	[.479]	[.154]
Female	0.506	0.479	0.499	[.369]	[.798]	[.490]
Birth order	1.75 (0.80)	1.70 (0.76)	1.77 (0.78)	[.335]	[.684]	[.180]
Put to the breast immediately	0.338	0.332	0.322	[.841]	[.685]	[.825]
Exclusively breastfed for 6-7 months	0.077	0.115	0.109	[.208]	[.276]	[.864]
Appropriately breastfed (at 0-23 months)	0.175	0.236	0.174	[.459]	[.999]	[.419]
Receiving 4+ food groups (at 6-23 months)	0.281	0.218	0.087	[.474]	[.014]	[.039]
Stunted (HAZ < -2)	0.700	0.660	0.683	[.213]	[.614]	[.489]
Wasted (WAZ < -2)	0.062	0.061	0.062	[.858]	[.932]	[.934]
Underweight (WHZ < -2)	0.359	0.340	0.320	[.542]	[.226]	[.585]
Malnourished (MUAC $<$ 125 mm)	0.057	0.044	0.061	[.310]	[.801]	[.231]
ASQ Communication score	36.41 (19.16)	39.27 (19.09)	37.16 (18.51)	[.106]	[.706]	[.218]
ASQ Gross Motor score	33.05 (19.24)	36.53 (19.19)	35.47 (18.19)	[.047]	[.199]	[.483]
New child						
Observations	685	719	744			
Months pregnant at baseline	5.48 (2.18)	5.46 (2.07)	5.29 (2.15)	[.816]	[.108]	[.170]

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The sample is also restricted to villages visited at midline – thus excluding those that could not be visited by survey teams because of security concerns. Each line corresponds to a variable. Columns (1) to (3) show mean (and standard deviation in parenthesis, if continuous) of the variable in each of the treatment groups. Columns (4) to (6) show *p*-values of the hypothesis that the mean of the variable is equal across C and T1, C and T2, and T1 and T2, respectively. These *p*-values are obtained by OLS regressions of the variable on dummies for the household residing in a T1 or T2 village, controlling for randomisation stratum (tranche) and clustering the standard errors at the village level. Variables are grouped based on whether they pertain the household as a whole, the woman, her husband, or the old child. For each group, the number of observation is reported. This number is not the same for all variables in each group due to missingness and skip patterns. The variables at the household level are defined as follows: *household size* is the number of people living in the household with common eating arrangements. *Progress out of Poverty Index (PPI)* is a measure of household wealth (Chen et al., 2008). *Equivalised daily per capita expenditure* is obtained by summing 42 items of food and non-food expenditure covering all areas, equivalising the amount using the OECD scale, and converting it from Nigerian Naira to PPP US dollars at the 2014 rate (99.4). *Woman and husband total monthly earnings* is obtained by summing the woman's and her husband's earnings across all work activities in the past 12 months, converting the amount to a monthly basis, and converting it to PPP US dollars as for expenditure. *Did not have enough food at some point in past year* is a dummy for households that report not having enough food to eat in the 4 seasons preceding the interview. The variables at the woman level are defined as follows: *Ever attended school* is a dummy for whether the woman has ever attended formal education. *Age (years)* is the woman's age in years. *Num. children aged 0-2* is the number of biological children of the woman residing in the household. *In polygamous marriage* is a dummy for the woman's marriage being polygamous. *Work activity in past year* is a dummy indicating whether the woman has undertaken any paid or unpaid work activity in the past year (excluding housework and childcare for the household she resides in. *Has a say on major HH purchases* is a dummy for whether the woman has any decision power on major household purchases (e.g. furniture). *Has a say on what food to buy* is a dummy for whether the woman has any decision power on what food is bought for household consumption. The variables at the husband level are defined as follows: *Ever attended school* is a dummy for whether the husband has ever attended formal education. *Can read and write* is a dummy for whether the husband is able to read and write any language. *Age (years)* is the husband's age in years. *Cultivated land in past year* is a dummy for whether the husband has cultivated land at any point in the past year. *Work activity in past year* is a dummy indicating whether the *Ever attended school* is a dummy for whether the husband has ever attended formal education; has undertaken any paid or unpaid work activity in the past year (excluding housework and childcare for the household he resides in. The variables at the old child level are defined as follows: *Age (months)* is the child's age in months. *Female* is a dummy indicating female gender. *Birth order* is the order of the child in the rest of the mother's biological children. *Put to the breast immediately* is a dummy for the child having been put to the breast in the first 30 minutes after birth. *Exclusively breastfed for at least 6m* is a dummy that takes value one only for children who have been exclusively breastfed for 6 months or more, and is not defined if the child is still being breastfed. *Appropriately breastfed* is a dummy indicating age-appropriate breastfeeding according to WHO guidelines (WHO, 2008), i.e. exclusive breastfeeding up to the age of 6 months and complementary breastfeeding from 6 to 23 months, and is not defined for children older that 23 months. *Receiving 4+ food groups* is a dummy indicating children who have received food from at least four among the following categories in the previous day: (i) Grains, roots and tubers; (ii) Legumes and nuts; (iii) Dairy products; (iv) Flesh foods; (v) Eggs; (vi) Vitamin-A rich fruits and veg.; (vii) Other fruits and veg; the indicator is not defined for children outside the 6-23 months age interval. *Stunted* is a dummy indicating children with height-for-age z-score (HAZ) under -2 SD. *Wasted* is a dummy indicating children with weight-for-age z-score (WAZ) under -2 SD. *Underweight* is a dummy indicating children with weight-for-height z-score (WHZ) under -2 SD. All z-scores are computed in accordance with WHO guidelines (WHO, 2009). *Malnourished* is a dummy indicating children with middle-upper-arm circumference (MUAC) lower than 125mm. *ASQ Communication score* and *ASQ Gross Motor score* are the raw scores from the communication and gross motor modules of the Ages and Stages Questionnaire, respectively. The variables at the new child level are defined as follows: *Months pregnant at baseline* is the number of month of pregnancy reported by the women respondents (if pregnant).

Table 2.3: Takeup of Unconditional Cash Transfer
Sample: Households with In Utero Pregnant Women at Baseline
Means, Standard Deviation in Parentheses, p-values in Brackets

	Control	**T1**	**T2**	**T1-T2 diff.**
	Mean (SD)	Mean (SD)	Mean (SD)	p-value
	(1)	(2)	(3)	(4)
Panel A: Cash transfer takeup				
Ever received grant	0.061	0.832	0.835	[.958]
If yes, still receiving grant at midline		0.910	0.900	[.689]
Panel B: Timing of first transfer				
Age of New Child at first payment (months)		−0.32 (4.32)	0.12 (4.81)	[.325]
During pregnancy		0.490	0.519	[.415]
1st trimester		0.031	0.019	[.286]
2nd trimester		0.140	0.108	[.111]
3rd trimester		0.319	0.392	[.018]
In same month of birth		0.150	0.128	[.301]
After birth		0.360	0.351	[.701]
Panel C: Cash transfer intensity				
Number of payments		19.81 (3.61)	19.65 (4.09)	[.931]
Frac. max payments (out of max. given child age)		0.82 (0.13)	0.83 (0.13)	[.275]
Total grant amount received (USD PPP)		687.90 (125.87)	681.62 (148.47)	[.939]
Randomisation strata				Yes
Observations	704	738	757	

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Each line corresponds to a variable. Columns (1) to (3) show mean (and standard deviation in parenthesis, if continuous) of the variable in each of the treatment groups; these are omitted for the control group apart from the overall take-up rate in the first row. Column (4) shows p-values of the hypothesis that the mean of the variable is equal across T1 and T2. These p-values are obtained by OLS regressions of the variable on dummies for the household residing in a T1 or T2 village, controlling for randomisation stratum (tranche) and clustering the standard errors at the village level. For each treatment arm, the number of observation is reported. This number is not the same for all variables in each group due to missingness and skip patterns. The variables are defined as follows: *Ever received grant* is a dummy for the woman reporting having aver participated in the CDGP cash transfer programme. *Still receiving grant at midline* is a dummy defined only for women who have ever received the grant, taking value 1 if they are still receiving the grant at the midline interview. *Age of New Child at first payment* is the age of the surveyed child born after the baseline survey at the time the first payment was received by the mother; it's computed by comparing the reported month of birth of the child (from the household questionnaire) with the reported month when the mother received her first CDGP payment (from the implementation dataset); negative ages imply payments received before birth. *During pregnancy*, *1st trimester*, ..., *After birth* are dummies indicating children who (according to the above age computation) have received the payment during pregnancy, in each of the 3 trimesters, in the same months as they were born, and after they had been born. *Number of payments* is the number of payments the mother received by the midline interview, according to the implementation dataset. *Frac. max payments* is the ratio between the number of payments received and the maximum number of payments to which the mother is entitled to; the denominator is computed by subtracting the month of first payment from the month when the child will turn 2 years old and the cash grant ends. *Total grant amount received* is the money amount of cash received up to the midline interview; this is calculated using data from the implementation dataset; it is then deflated to August 2014 using the monthly rural CPI index published by the Central Bank of Nigeria CITE, and finally converted using the PPP US dollar rate for August 2014.

Table 2.4: Exposure to Low- and High-Intensity Information
Sample: Households with In Utero Pregnant Women at Baseline
Means, p-values in Brackets

	Women						**Husbands**						**Women vs. Husbands**		
	C	**T1**	**T2**	**pvalues of difference**			**C**	**T1**	**T2**	**pvalues of difference**			**pvalues of difference**		
	Mean	Mean	Mean	T1-C	T2-C	T2-T1	Mean	Mean	Mean	T1-C	T2-C	T2-T1	C	T1	T2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Low-intensity channels															
None	0.408	0.095	0.077	[.000]	[.000]	[.440]	0.343	0.151	0.197	[.000]	[.000]	[.138]	[.026]	[.015]	[.000]
At least one	0.592	0.905	0.923	[.000]	[.000]	[.440]	0.657	0.849	0.803	[.000]	[.000]	[.138]	[.026]	[.015]	[.000]
All	0.006	0.160	0.128	[.000]	[.000]	[.313]	0.005	0.037	0.034	[.001]	[.002]	[.887]	[.711]	[.000]	[.000]
Panel B: High-intensity channels															
None	0.935	0.457	0.332	[.000]	[.000]	[.016]	0.942	0.851	0.859	[.000]	[.000]	[.836]	[.668]	[.000]	[.000]
All	0.007	0.107	0.151	[.000]	[.000]	[.081]	0.005	0.028	0.030	[.051]	[.003]	[.839]	[.537]	[.000]	[.000]
Support group		0.520	0.655			[.008]		0.122	0.096			[.337]		[.000]	[.000]
Says 1:1 counselling available in village		0.374	0.448			[.174]		0.296	0.358			[.218]		[.009]	[.002]
If yes, tried to obtain 1:1 counselling		0.406	0.375			[.555]		0.217	0.228			[.778]		[.001]	[.001]
If yes, obtained 1:1 counselling		0.857	0.976			[.001]		0.857	0.921			[.482]		[.947]	[.282]
Randomisation strata				Yes	Yes	Yes				Yes	Yes	Yes	Yes	Yes	Yes
Observations	704	738	756				417	436	467						

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Each line corresponds to a variable. Columns (1) to (3) show mean of sampled women's exposure to BCC channels in each of the treatment groups. Columns (4) to (6) shows p-values of the hypothesis that the mean of the variable is equal across T1 and T2. These p-values are obtained by OLS regressions of the variable on dummies for the household residing in a T1 or T2 village, controlling for randomisation stratum (tranche) and clustering the standard errors at the village level. For each treatment arm, the number of observation is reported at the bottom of the table. This number is not the same for all variables in each group due to missingness and skip patterns. Columns (7) to (12) replicate the same analysis for husbands of the sampled women. Columns (13) to (15) report p-values for tests of equality between women's and husbands' exposure (when both are surveyed); these p-values are obtained by OLS regressions of each variables on dummies for treatment arm and respondent – woman vs. husband – and their interactions, controlling for randomisation stratum (tranche), and clustering the standard errors at the village level. Low-intensity channels are posters, radio programmes/ads, health talks, food demonstrations, and voice messages. Men are not surveyed about receiving voice messages, so their maximum number of low-intensity channels is four – while it's 5 for women. High-intensity channels are support groups and one-to-one counselling. *None* of the low-intensity channels is a dummy for the respondent reporting not having been exposed to any of the channels. *At least one* is a dummy for the respondent reporting having been exposed to at least one of the five low-intensity channels (four for men). *All* is a dummy for being exposed to all of the channels. Participation to one-to-one counselling is surveyed sequentially: first, respondents are asked if such activity is available in their village; in case of a positive answer, they are asked whether they ever tried to access it; in case of a positive answer, they are asked if they were able to obtain it.

Table 2.5: Number and timing of births

Sample: Pregnant Women at Baseline

ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean (1)	**Effect** (2)	**Effect** (3)
A. Pregnant at baseline			
Had any live birth between BL and ML	93.72	1.63* (0.91)	0.86 (1.05)
Spacing after last child (months)	33.42 (12.78)	0.03 (0.55)	−0.16 (0.61)
Had any child born between BL and ML who died	12.94	0.57 (1.25)	0.11 (1.26)
Number of children born between BL and ML who died	0.14 (0.37)	0.00 (0.01)	−0.00 (0.01)
Month of birth of New Child		0.54** (0.27)	0.58* (0.32)
Woman is pregnant at ML	36.77	−1.21 (1.84)	−1.31 (1.92)
B. Main estimation sample			
Spacing after last child (months)	33.60 (11.84)	1.11** (0.51)	1.00* (0.55)
Had any child born between BL and ML who died	0.84	0.06 (0.43)	0.44 (0.52)
Number of children born between BL and ML who died	0.01 (0.09)	0.00 (0.00)	0.01 (0.01)
Month of birth of New Child		0.55*** (0.11)	0.45*** (0.11)
Woman is pregnant at ML	40.00	−0.95 (2.21)	0.75 (2.39)
Week of BL/ML interview dummies			✓

Notes: The sample for Panel A is all women who reported being pregnant at baseline, regardless whether they had any live birth or whether the child surveyed at midline was in utero at baseline or not. The sample for Panel B is restricted to the main sample used for subsequent estimation, i.e. households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Col. (3) adds dummies for the week when the baseline and midline interviews took place. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.6: New child anthropometrics

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	Raw Effect	Age-adjusted effect
	(1)	(2)	(3)
Continuous indicators			
Height-for-age (HAZ)	-2.72 (1.12)	0.21^{***} (0.06)	0.12^{**} (0.06)
Weight-for-age (WAZ)	-1.86 (1.08)	0.05 (0.06)	0.02 (0.06)
Weight-for-height (WHZ)	-0.64 (1.05)	-0.10^{*} (0.06)	-0.07 (0.06)
Middle Upp. Arm Circumf. (MUAC)	136.47 (11.77)	-0.06 (0.65)	0.12 (0.65)
Thresholds			
Stunted (HAZ$<$-2)	73.47	-4.82^{**} (2.45)	-1.76 (2.43)
Underweight (WAZ$<$-2)	42.51	-0.23 (2.61)	1.17 (2.62)
Wasted (WHZ$<$-2)	9.99	3.42^{**} (1.49)	2.76^{*} (1.53)
Malnourished (MUAC$<$125mm)	14.00	0.22 (1.78)	-0.07 (1.79)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Height-for-age, weight-for-age, and weight-for-height indicators are computed with reference to the WHO growth standards (WHO, 2009). Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.7: New child communication and motor skills

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	Raw Effect	Age-adjusted effect	Age-adjusted effect
	(1)	(2)	(3)	(4)
Communication skills (Z)	-0.00	0.08	0.11^{*}	0.11^{*}
	(1.00)	(0.06)	(0.06)	(0.06)
Communication skills in normal range	30.76	5.74^{**}	5.73^{**}	5.18^{**}
		(2.36)	(2.38)	(2.36)
Gross motor skills (Z)	0.00	0.09	0.07	0.03
	(1.00)	(0.06)	(0.06)	(0.06)
Gross motor skills in normal range	40.73	4.14	2.91	2.25
		(3.01)	(3.06)	(2.96)
Age group dummies			✓	
ASQ age bands + interaction				✓

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Communication and motor skills Z scores are obtained by standardising raw scores using the mean and standard deviation in the control group at midline. 'Normal ranges' are binary indicators that take value 1 if the score falls in the normal category based on the reference population. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made in the same way as Table 2.6, i.e. using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Col. (4) conducts a different age adjustment, using the same age bands used to decide which ASQ module is administered (in months): 13-14, 15-16, 17-18, 19-20, 21-22, 23-25.5, 25.5-28. It also allows the treatment effect to vary at each age group by interacting it with the age dummies. The effect displayed in the table is the sample-weighted average of age-specific treatment effects. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.8: Child Health

Sample: In Utero New Child, Older Sibling of In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	New Child		Old Child		New-Old Diff.
	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)
Number of vaccinations	1.97 (1.48)	0.51^{***} (0.11)			.
Given deworming medication in past 6 months	17.12	9.65^{***} (2.19)	20.30	10.47^{***} (2.92)	0.833
Had illness/injury in past 30 days	72.47	-7.02^{***} (2.59)	67.49	-8.02^{***} (3.10)	0.912
Had diarrhoea in past 2 weeks	38.06	-7.11^{***} (2.32)	20.88	-6.32^{**} (2.45)	0.557

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. *Number of vaccinations* is an index that considers the number of vaccinations the child received among the following: BCG, polio, DPT, hepatitis B, yellow fever, and measles. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the Old Child (aged 0-5 at baseline), if one was surveyed. Column (5) shows the p-value for the hypothesis that the estimated effects are equal across Old and New children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.9: Parental Knowledge of IYCF practices

Sample: Households with In Utero Pregnant Women at Baseline
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	Woman		Husband		Woman-Husband Diff.
	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)
Knowledge Index (using †)	0.00 (1.00)	1.07*** (0.10)	−0.00 (1.00)	0.74*** (0.09)	0.000
Prenatal					
Would advise to seek antenatal care even if healthy†	68.35	7.75*** (2.12)	72.89	5.06** (2.21)	0.219
Perinatal					
Colostrum is good for the baby†	65.01	20.08*** (2.69)	57.16	12.57*** (3.17)	0.033
Best to start breastfeeding immediately†	17.42	27.44*** (2.97)	18.54	12.66*** (3.02)	0.000
Best place to give birth is health facility†	15.09	12.08*** (2.93)	19.46	10.90*** (3.56)	0.571
Baby should not receive other liquids on first day†	46.49	23.68*** (2.98)	49.72	21.10*** (3.60)	0.426
Postnatal					
Not ok to give water to baby under 6 months when hot†	7.55	40.23*** (3.54)	12.36	26.14*** (3.14)	0.000
Best to breastfeed exclusively for 6-7 months†	20.27	30.07*** (3.37)	0.00		

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ANCOVA ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. The index variable is computed using the methodology in Anderson (2008), and is standardised to have mean zero and variance one in the control group at baseline. The weights for the index are computed using the entire sample at baseline. The last indicator (*Best to breastfeed exclusively for 6-7 months*) has zero prevalence among husbands, so the effects cannot be computed; it is also excluded from the computation of the husbands' knowledge indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at baseline, for the surveyed women. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the husbands of the surveyed women. Column (5) shows the p-value for the hypothesis that the estimated effects are equal across women and husbands. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.10: New Child antenatal care and breastfeeding practices
Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	Effect
	(1)	(2)
New Child practices index (using †)	-0.00 (1.00)	0.55*** (0.08)
Prenatal		
Received antenatal care†	60.96	8.96** (3.86)
Perinatal		
Fed colostrum in the first hour†	37.97	28.86*** (3.17)
Put to the breast immediately†	44.32	25.97*** (3.13)
Born at health facility†	12.78	5.48** (2.14)
Postnatal		
Exclusively breastfed for 6-7 months†	11.53	30.13*** (2.97)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. The index variable is computed using the methodology in Anderson (2008), and is standardised to have mean zero and variance one in the control group. The weights for the index are computed using the entire sample at midline. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.11: Child Dietary Diversity

Sample: In Utero New Child, Older Sibling of In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	New Child		Old Child		New-Old Diff.
	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)
Dietary diversity index (using †)	3.28 (1.13)	0.35*** (0.07)	3.53 (1.00)	0.26*** (0.06)	0.204
Food groups					
Grains, tubers, roots†	98.46	0.08 (0.65)	99.16	0.14 (0.50)	0.557
Fruit and vegetables	85.53	1.37 (1.80)	94.13	−0.57 (1.26)	0.390
Dark green leafy vegetables	43.26	−8.80*** (2.46)	50.31	−8.51** (3.32)	0.411
Vit-A rich fruit and vegetables†	64.61	8.03*** (2.34)	72.54	6.33** (2.83)	0.574
Other fruit and vegetables†	45.22	8.10*** (2.83)	53.67	6.15* (3.17)	0.467
Nuts, beans, and seeds†	60.81	4.59* (2.48)	66.67	1.70 (3.07)	0.449
Animal-source foods	38.76	15.56*** (2.65)	41.51	10.98*** (3.10)	0.055
Flesh foods (meat, organ meat, fish)†	15.17	6.52*** (2.05)	17.19	6.62*** (2.43)	0.795
Eggs†	0.70	1.06** (0.44)	0.42	0.41 (0.34)	0.188
Milk, cheese, yogurt†	28.23	14.06*** (2.53)	28.72	10.12*** (2.87)	0.049

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. The diet diversity index and the food groups are obtained from a 24-h food recall module administered to the child's mother or main carer. Each meal consumed in the day before the interview from waking up to bedtime is recorded, and each ingredient is coded into categories. The index uses slightly different food groups as the one presented below, according to WHO (2008). It is derived by summing the number of food groups the child has received, among the following categories: (i) Grains, roots and tubers; (ii) Legumes and nuts; (iii) Dairy products; (iv) Flesh foods; (v) Eggs; (vi) Vitamin-A rich fruits and veg.; (vii) Other fruits and vegetables. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the Old Child (aged 0-5 at baseline), if one was surveyed. Column (5) shows the p-value for the hypothesis that the estimated effects are equal across Old and New children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child and the child's age in months. Significance: * (10%); ** (5%); *** (1%).

Table 2.12: Household Expenditures and Savings
Sample: Households with In Utero Pregnant Women at Baseline
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean (1)	Effect (2)
Panel A: *Monthly expenditure (PPP USD)*		
Total	201.34 (239.48)	27.82* (14.38)
Food	85.30 (124.71)	20.26** (8.93)
Non-food	131.48 (157.87)	8.89 (10.36)
Durables	4.30 (16.22)	0.90 (0.80)
Panel B: *Equivalised monthly expenditure (PPP USD)*		
Total	45.64 (52.23)	5.02 (3.24)
Food	19.67 (28.58)	4.98*** (1.90)
Non-food	29.47 (33.73)	0.70 (2.39)
Durables	0.93 (3.86)	0.16 (0.19)
Panel C: *Borrowing and saving*		
Any household member borrowing money	24.22	-1.85 (2.47)
Tot. value of borrowing (PPP USD)	32.60 (127.42)	-15.23 (9.67)
Any household member saving money (incl. in-kind)	64.07	2.81 (2.26)
Tot. value of savings (incl. in-kind, PPP USD)	244.90 (631.49)	-16.12 (68.67)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ANCOVA ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Food expenditure is obtained using a 7-day expenditure recall of 13 food items: (i) Foods made from grains; (ii) Dark green leafy vegetables; (iii) Potatoes and roots; (iv) Other vegetables; (v) Fruit; (vi) Nuts and beans; (vii) Meat and eggs; (viii) Fish; (ix) Milk, cheese, and yogurt; (x) Oils and butter; (xi) Condiments for flavour; (xii) Sugary foods and sweets; (xiii) Drinks. Non-food expenditure is obtained using: a 7-day expenditure recall of consumables (e.g. matches, fuel), a 30-day recall (e.g. toiletries, utensils, household items, health expenditure), and a 12-month recall of major expenses (e.g. ceremony costs, school fees, remittances). Expenditure on durables is obtained using a 12-month recall of expenditure on assets the household owns (e.g. TV set, wheelbarrow, mattress). All expenditure, borrowing, and savings amounts are trimmed of their top percentile to reduce the impact of potential outliers. Expenditures in the Panel A are deflated to August 2014 Naira amounts using the Nigeria rural CPI (Central Bank of Nigeria, 2017), and then converted to USD using the 2014 PPP exchange rate (World Bank, 2017). In Panel B, expenditures are equivalised using OECD adult-equivalent scales (OECD, 1982). Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at baseline. Col. (2) shows the ANCOVA ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the new child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.13: Food Expenditure and Food Security

Sample: Households with In Utero Pregnant Women at Baseline
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean (1)	Effect (2)
Panel A: *Monthly food expenditure (PPP USD)*		
Grains, tubers, roots	30.37 (63.66)	4.81 (4.37)
Fruits and vegetables	7.97 (14.32)	1.57 (1.15)
Dark green leafy vegetables	2.15 (4.20)	0.54^{*} (0.30)
Other fruit and vegetables	5.65 (11.48)	0.43 (0.95)
Nuts, beans, and seeds	3.52 (9.14)	0.42 (0.59)
Animal-source foods	22.29 (36.10)	7.54^{***} (2.26)
Meat and eggs	14.74 (28.91)	4.09^{**} (1.78)
Fish	4.21 (9.40)	1.63^{***} (0.61)
Milk, cheese, yogurt	2.81 (6.79)	1.04^{*} (0.56)
Other (condiments, oils, drinks)	13.80 (20.65)	0.16 (1.75)
Panel B: *Food security*		
Had not enough food in past 12 months	13.76	-8.59^{***} (2.43)
Reduced number of meals in past 30 days	17.70	-7.97^{***} (2.25)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ANCOVA ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. All expenditure categories are derived from 7-day recalls of expenditure, trimmed of their top centile to reduce the impact of potential outliers. They are also deflated to August 2014 Naira amounts using the Nigeria rural CPI (Central Bank of Nigeria, 2017), and then converted to USD using the 2014 PPP exchange rate (World Bank, 2017). All expenditures are converted into monthly amounts. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at baseline. Col. (2) shows the ANCOVA ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the new child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.14: Woman and Husband Labour

Sample: Households with In Utero Pregnant Women at Baseline
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	Woman		Husband		Woman-Husband Diff.
	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)
Extensive margin (past 12 months)					
Any work activity	75.99	7.60^{***} (1.97)	99.71	0.24 (0.22)	0.001
Cultivated land	5.41	-0.72 (1.39)	96.41	-1.13 (0.94)	0.813
Reared animals	40.91	4.13 (3.39)	33.00	3.72 (2.86)	0.821
Had self-employed activity	65.48	7.15^{***} (2.38)	61.41	-3.03 (2.81)	0.003
Intensive margin					
Number of activities in past 12 months	1.17 (0.85)	0.11^{**} (0.05)	2.08 (0.70)	0.03 (0.04)	0.093
Days/week usually worked at highest-earning act.	3.20 (3.05)	0.32^{*} (0.17)	4.23 (2.59)	0.09 (0.14)	0.386
Earnings					
Sold any crops	1.57	0.81 (0.82)	50.50	-3.17 (2.47)	0.410
Monthly revenue from crop sales (PPP USD)	0.10 (1.39)	0.14 (0.09)	28.97 (54.56)	-0.75 (2.76)	0.877
Monthly earnings from work activities (PPP USD)	25.04 (38.66)	4.47^{**} (2.00)	141.42 (283.72)	10.52 (15.98)	0.617

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Work activities are defined as any paid or unpaid work, either self-employed or salaried, excluding housework and childcare. We define self-employed activities as ones where payments are received directly from the client/customer (e.g. hairdresser working in her own shop, or labouring on one's own household land) rather than from an employer (e.g. receiving a salary to work as an hairdresser in someone else's shop, or labouring on someone else's land for a payment). Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the surveyed women. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the husbands of the surveyed women. Column (5) shows the p-value for the hypothesis that the estimated effects are equal across women and husbands. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the new child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.15: Food Engel Curves

Sample: Households with In Utero Pregnant Women at Baseline
Gelbach Decomposition Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	ITT		Engel Curve	
	(1)	(2)		(3)	(4)
Panel A: Overall food share					
All women	48.97 (20.29)				
Treated		2.10 (1.34)	α_2	3.40** (1.40)	3.23** (1.34)
(log) nondur. exp.			α_1	−7.95*** (2.01)	−5.79 (3.78)
Treated ∗ (log) nondur. exp.			α_3		−3.17 (4.78)
High bargaining power at BL	50.64 (20.58)				
Treated		1.55 (1.59)	α_2	3.32* (1.73)	2.65 (1.68)
(log) nondur. exp.			α_1	−8.98*** (3.23)	−1.26 (4.52)
Treated ∗ (log) nondur. exp.			α_3		−12.99* (6.70)
Low bargaining power at BL	47.23 (19.89)				
Treated		2.55 (1.75)	α_2	3.50* (1.80)	3.93** (1.83)
(log) nondur. exp.			α_1	−7.09*** (2.69)	−14.56* (8.34)
Treated ∗ (log) nondur. exp.			α_3		9.71 (9.10)
Panel B: Food shares by group					
Share Animal-source foods	23.87 (20.97)				
Treated		3.93*** (1.10)	α_2	2.08 (1.44)	3.55 (3.40)
(log) food exp.			α_1	6.58*** (2.45)	8.35** (3.32)
Treated ∗ (log) food exp.			α_3		−2.78 (5.09)
Share Meat and eggs	14.79 (18.21)				
Treated		2.19** (0.96)	α_2	0.06 (1.21)	0.78 (2.62)
(log) food exp.			α_1	7.38*** (2.02)	8.24*** (2.29)
Treated ∗ (log) food exp.			α_3		−1.35 (3.99)
Share Fish	4.65 (7.73)				
Treated		1.22** (0.48)	α_2	1.35** (0.53)	2.08* (1.24)
(log) food exp.			α_1	−0.49 (0.90)	0.36 (1.19)
Treated ∗ (log) food exp.			α_3		−1.38 (1.87)
Share Milk, cheese, yogurt	4.48 (9.87)				
Treated		0.28 (0.58)	α_2	0.39 (0.58)	0.40 (1.75)
(log) food exp.			α_1	−0.42 (0.94)	−0.41 (2.21)
Treated ∗ (log) food exp.			α_3		−0.02 (2.73)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports instrumental variable estimates of Engel curve parameters. The first set of results considers the share of food expenditure over total expenditure on nondurables. Subsequent results instead consider the share of expenditure on each food category over total food expenditure. All shares are reported in percentage points. We demean total expenditure on durables and total food expenditure to approximate effects at the mean. For each food share, Col. (1) shows the mean and SD in the control group at midline. Col. (2) shows the ITT effect in the restricted model on the budget share. Col. (3) shows two-stage least squares estimates of Engel curve parameters, obtained by regressing the budget share on its denominator (nondurable expenditure or food expenditure) and the treatment indicator. This corresponds to (2.5.1) with α_3 restricted to be zero. Col. (4) allows the slope shift parameter α_3 to be different from zero. We instrument total nondurable expenditure using household wealth, as measured by the PPI index (Schreiner, 2015). All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. Significance: * (10%); ** (5%); *** (1%).

2.8 Figures

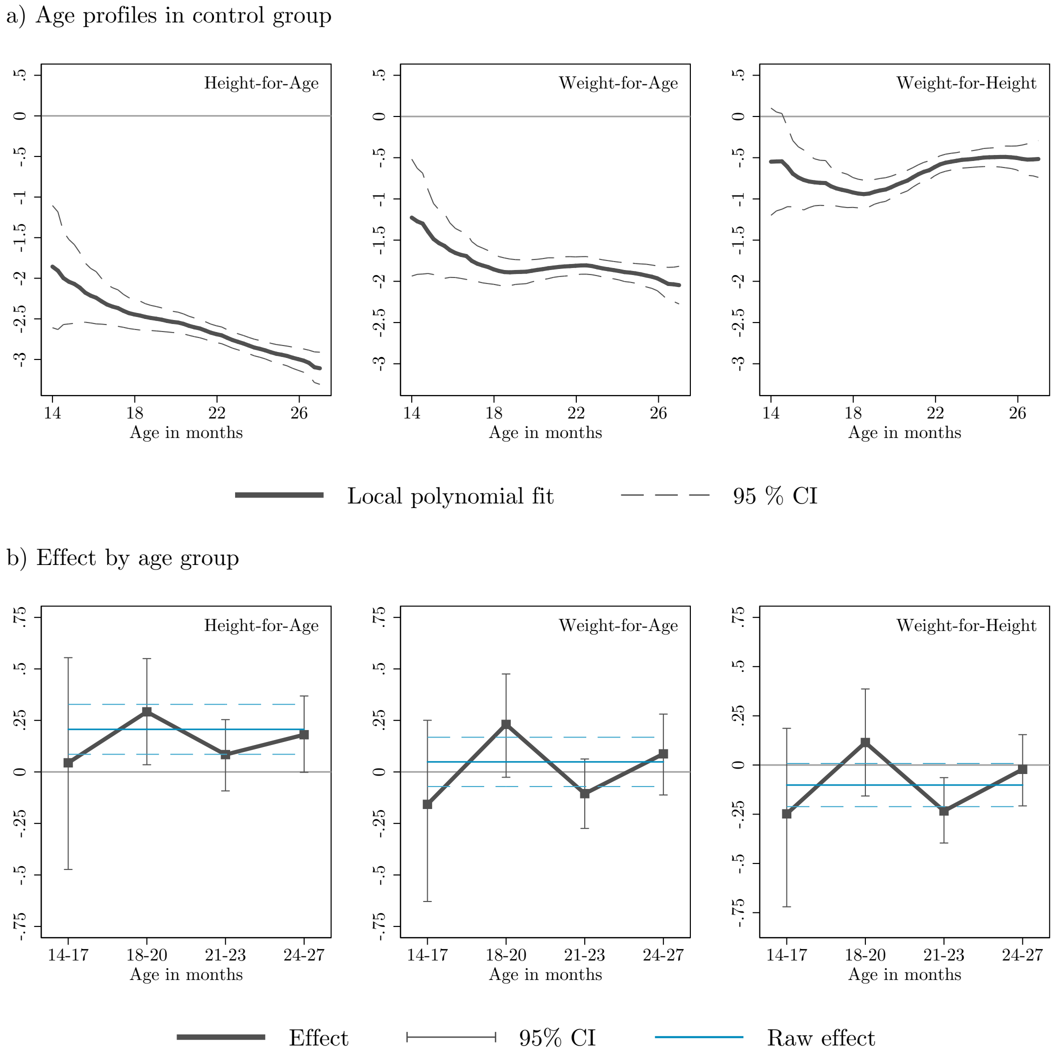

Figure 2.1: New Child anthropometrics - raw vs. control function estimates
Sample: In Utero New Child

Notes: The main sample for this graph is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Panel a) shows smoothed age profiles of mean anthropometric z-scores in the control group. The profiles are obtained using a local mean kernel smoother. Panel b) displays effects on anthropometric z-scores by age group. These correspond to the ones used to compute the average effects in col. (7) of Table 2.A8. Standard errors for the confidence intervals are obtained by bootstrap with 1,000 repetitions. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

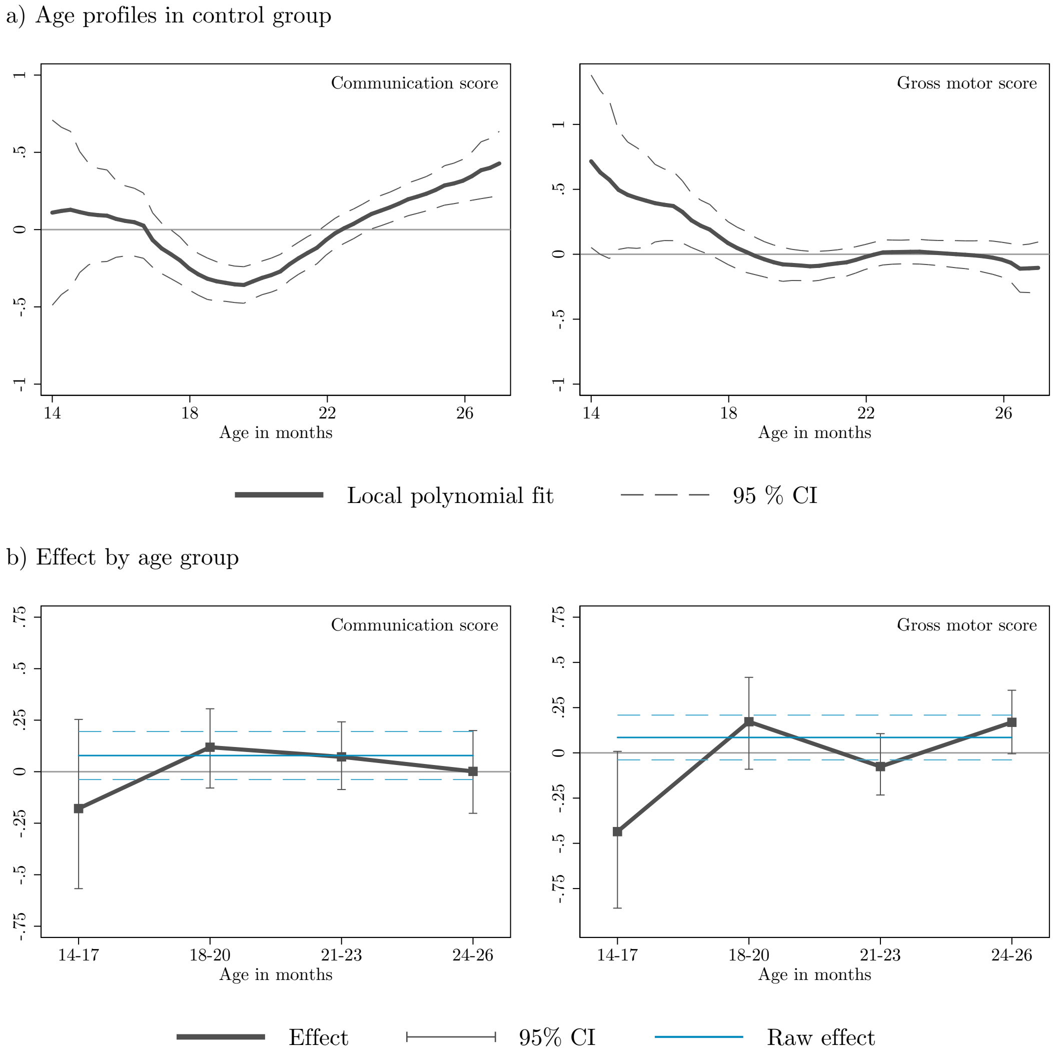

Figure 2.3: New Child communication and motor skills - raw vs. control function estimates
Sample: In Utero New Child

Notes: The main sample for this graph is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Panel a) shows smoothed age profiles of mean ASQ scores (internally standardised to have mean zero and variance one) in the control group. The profiles are obtained using a local mean kernel smoother. Panel b) displays effects on ASQ scores by age group. These correspond to the ones used to compute the average effects in col. (7) of Table 2.A13. Standard errors for the confidence intervals are obtained by bootstrap with 1,000 repetitions. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

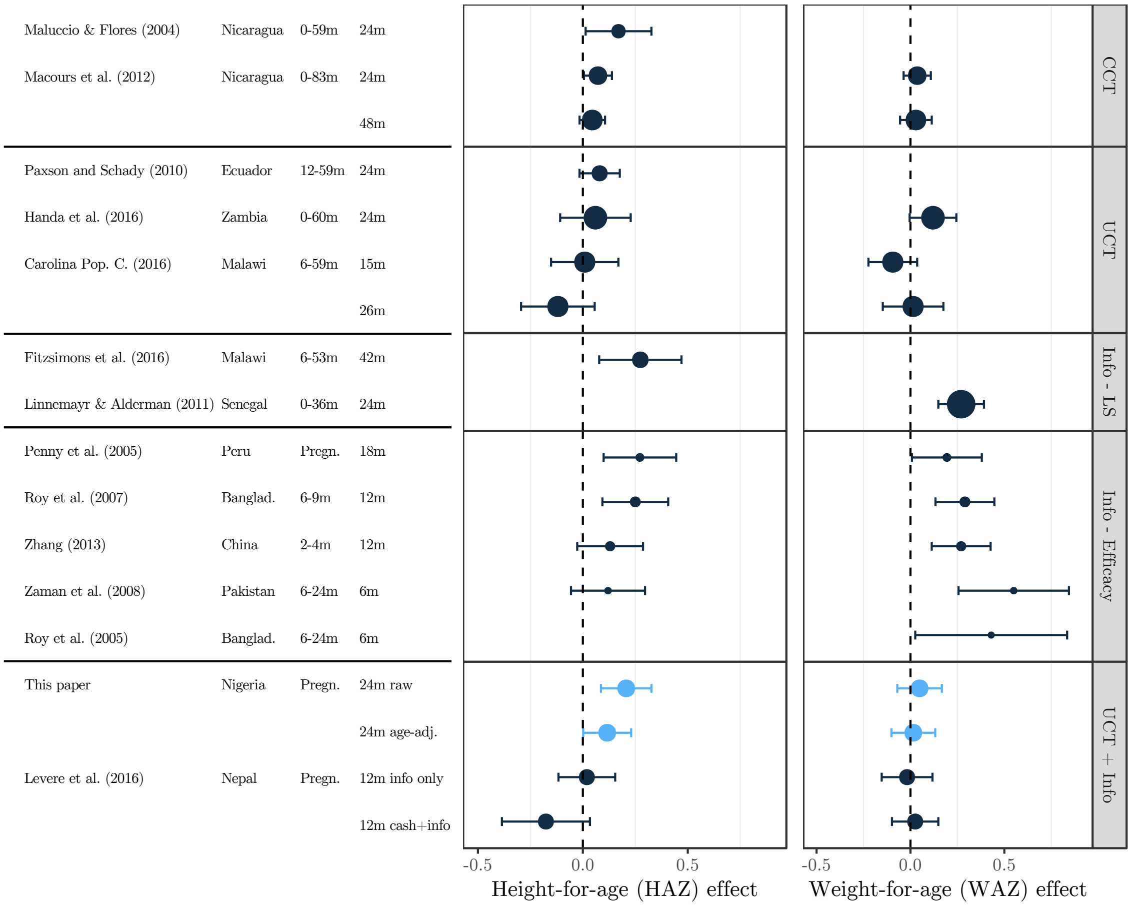

Figure 2.5: Impacts of RCT Cash Transfers on HAZ and WAZ Outcomes

Notes: The figure graphically depicts estimates of effects on height-for-age (HAZ) and weight-for-age (WAZ) from other studies. We restrict our attention to randomised interventions taking place in low- and middle-income countries where either HAZ or WAZ effects are estimated. The first column reports the authors and year of the study. The interventions are divided into five panels corresponding to as many categories: conditional cash transfers (CCT); unconditional cash transfers (UCT); information/education-only interventions implemented at a relatively large scale, or 'effectiveness' studies (Info - LS); information/education-only interventions on a smaller scale and closer to ideal conditions, or 'efficacy' studies (Info - Efficacy); and unconditional cash transfers with an added information/education component (UCT + Info). The second column reports the country where the intervention took place. The third column shows the age range at which children were initially exposed to the intervention. 'Pregn.' denotes interventions starting during pregnancy. The fourth column reports the follow-up horizon for the evaluation. The last two columns graphically depict point estimates and 95% confidence intervals for the estimated effects on HAZ and WAZ. The size of the dot is proportional to the sample size.

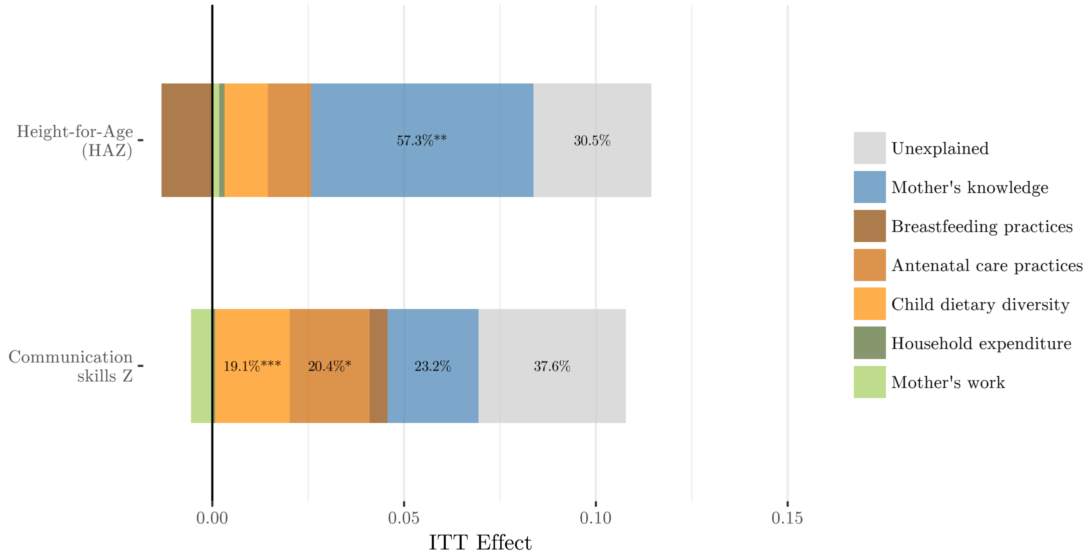

Figure 2.6: Mediation analysis
Sample: In Utero New Child

Notes: The main sample for this figure is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The figure reports decompositions of the ITT effect according to the methodology in Gelbach (2016). The size of each horizontal bar corresponds to the total ITT effect, and each section of the bar is the part that can be explained by the relative group of mediators. Percentages on the largest sections indicate the proportion of the total size of the bar explained by each mediator group. The mediator groups are defined as follows: *Mother's knowledge* [Index of maternal knowledge about ECD practices, see Table 2.9] *Breastfeeding practices* related to the new child [Fed colostrum in the first hour, put to the breast immediately, exclusively breastfed for 6-7 months], *Antenatal care practices* related to the new child [Received antenatal care, born at health facility], *Child dietary diversity* [Dietary diversity index, see Table 2.11], *Total HH expenditure*, *Mother's work* [Any work activity in past 12 months, number of activities]. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, and the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

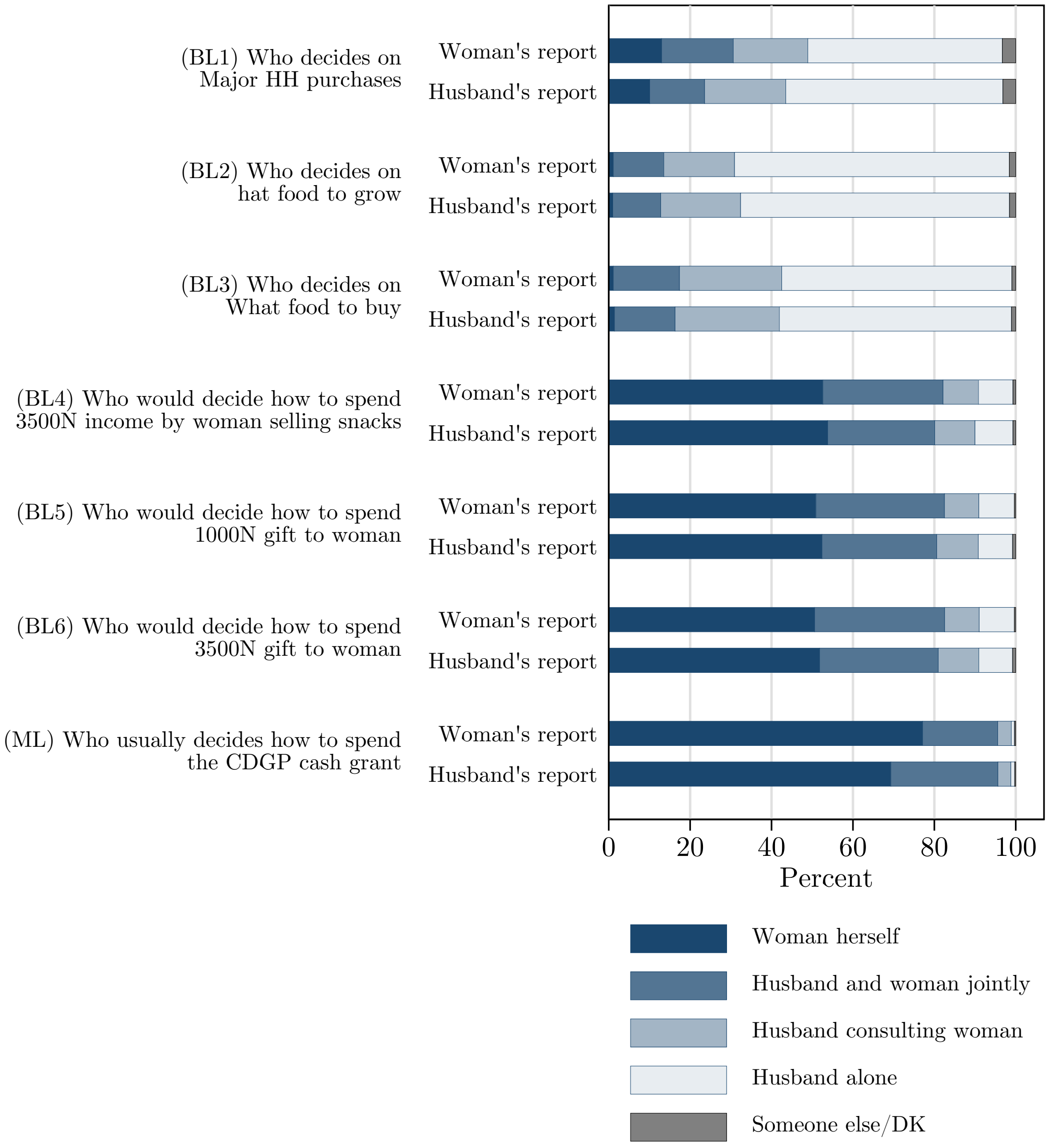

Figure 2.8: Intra-household decision making

Notes: The main sample for this figure is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The figure shows women and their husbands' responses to questions about bargaining power. The following questions were asked at baseline: (BL1) Who usually makes decisions about making major household purchases? These purchases are non-food related such as mattress and furniture. (BL2) Who usually makes decisions about what food to grow for the household to eat? (BL3) Who usually makes decisions about what food to buy? (BL4) Suppose you were to make 3500 Naira in 30 days selling snacks. Who do you think would decide how this money was used? (BL5) Now suppose you were to be given a regular monthly gift of 1000 Naira, and that this money is given only to you and not to any other household member. Who do you think would decide how this money was used? (BL6) Suppose you were to be given a regular monthly gift of 3500 Naira, and that this money is given only to you and not to any other household member. Who do you think would decide how this money was used? The following question was asked to women who had been recipients of the cash grant at midline: (ML) Who in the household usually decides how the payments are spent? The shading of each area of the bars represents the fraction of respondents giving each answer.

2.9 Appendix

2.9.1 Detailed sampling strategy

Our sampling procedure is outlined in detail here:

1. Take list of all villages in the five LGAs where the CDGP is operating
2. Drop the 15 villages used in the CDGP pilot
3. Drop villages with less than 150 households
4. Randomly sample 210 villages
5. Select one traditional ward per village using probability proportional to size within village
6. Select one replacement traditional ward per village to be used only in the case where the original sampled traditional ward is not accessible for security reasons
7. Send listing team to selected traditional wards
8. Replace traditional ward if listing teams find security problems when they arrive
9. Team to meet with traditional leaders and estimate size of traditional ward
10. If traditional ward contains:
    a) 0-200 households, list whole traditional ward
    b) 200-400 households, divide into two roughly equal sized parts
    c) 400-800 households, divide into four roughly equal sized parts
    d) 800+ households divide into eight roughly equal sized parts
11. If the situation of 10b, 10c, or 10d arises, randomly select one 'part' using random number table and list all households in randomly selected 'part'
12. The supervisor counts number of households that have been listed
13. If listing contains 0-100 households then:
    - 'Mapper' must make a list of all neighbouring, contiguous traditional wards
    - Randomly select an additional traditional ward using random number table
    - List this traditional ward following steps 8, 9 as 10, as stated above
14. If listing contains 100+ households continue to next step
15. Sample 26 households per village. If there are more than 26 households with at least one pregnant woman in the village, use simple random sampling to sample 26 households with at least one pregnant woman. If there are less than 26 households with at least one pregnant woman in the village, sample all households with at least one pregnant woman and make up the rest of the sample in that village with households containing at least one woman determined to be 'likely to become pregnant' (as estimated by the prediction model outlined in Section 2.3).

2.9.2 Mediation analysis

Suppose the following *base* model for the effect of an intervention T on outcome Y has been estimated by OLS:

$$Y = \alpha + \beta_1 T + \varepsilon \tag{2.9.1}$$

Also assume that, like in CDGP, the parameter β_1 can be given a causal interpretation – e.g. an ITT effect. The model can be thought of as being implicitly conditional on covariates $\boldsymbol{X}$, omitted for clarity. The interest lies in decomposing the ITT into components that are attributable to a set of k_m additional variables (mediators) $\boldsymbol{M}$.

Consider the *full* model:

$$Y = \alpha + \beta_1 T + \boldsymbol{M}\beta_2 + \varepsilon \tag{2.9.2}$$

The population value for the OLS estimate of β_1 in the base model (2.9.1) is:

$$\beta_1^{\text{base}} = \beta_1 + \boldsymbol{\Gamma}\beta_2 = \beta_1 + \delta \tag{2.9.3}$$

where Γ is the matrix of coefficients from a linear projection of $\boldsymbol{M}$ on T:

$$\boldsymbol{M} = T\boldsymbol{\Gamma} + \boldsymbol{u} \tag{2.9.4}$$

The decomposition in (2.9.3) gives the well-known omitted variables bias in the population resulting from excluding $\boldsymbol{M}$ from the base model.

The projection in (2.9.4) is split among each mediator variable as:

$$M^1 = \gamma_0^1 + \gamma_1^1 T + u^1$$

$$\dots$$

$$M^k = \gamma_0^{k_m} + \gamma_1^{k_m} T + u^{k_m}$$

Thus, a straightforward decomposition of the bias term across mediators is:

$$\delta = \beta_1^{\text{base}} - \beta_1 = \gamma_1^1 \beta_2^1 + \dots + \gamma_1^{k_m} \beta_2^{k_m} = \sum_{j=1}^{k_m} \gamma_1^j \beta_2^j$$

where $\delta^j = \gamma_1^j \beta_2^j$ is the component due to each mediator.

This suggests a simple algorithm to perform the decomposition (Gelbach, 2016):

1. Estimate the full model in (2.9.2) to get $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_1, \hat{\beta}_2\}$
2. Estimate a set of auxiliary models with each mediator acting as a dependent variable, and retrieve coefficients $\hat{\boldsymbol{\Gamma}}$
3. Get $\hat{\delta}^k$ by computing the product $\hat{\gamma}_1^k \hat{\beta}_2^k$

The simple algorithm can be extended to accommodate decomposition across groups of variables instead of single variables (Gelbach, 2016, p. 523). Standard errors for $\hat{\delta}^k$ are computed as detailed in Appendix B of Gelbach (2016).

2.9.3 Appendix tables

Table 2.A1: Setting

Sample: Households with In Utero Pregnant Women at Baseline
Means, Standard Deviations in Parentheses

	All Mean (SD) (1)	**Jigawa** Mean (SD) (2)	**Zamfara** Mean (SD) (3)
Livelihoods			
Equivalised daily per capita exp. (USD PPP)	1.65 (2.12)	1.71 (2.19)	1.61 (2.07)
Under poverty line	0.718	0.697	0.733
Woman and husband total monthly earnings (USD PPP)	170.60 (337.96)	158.90 (309.48)	179.00 (356.87)
Owns mobile phone	0.597	0.619	0.580
Did not have enough food at some point in past year	0.144	0.181	0.117
Woman's activities			
Cultivated land	0.041	0.072	0.019
Any Work activity	0.719	0.702	0.731
Ethnicity and religion			
Hausa ethnicity	0.908	0.832	0.963
In polygamous marriage	0.482	0.407	0.536
Woman's decision making power			
Has a say on major HH purchases	0.490	0.418	0.542
Has a say on what food to buy	0.425	0.362	0.471
Woman's advice seeking on IYCF			
Would go to mother	0.282	0.315	0.260
Would go to mother-in-law	0.147	0.117	0.168
Would go to trained health worker	0.227	0.282	0.188
Child nutrition practices			
Put to the breast immediately	0.331	0.228	0.407
Exclusively breastfed for 6-7 months	0.101	0.094	0.106
Appropriately breastfed (at 0-23 months)	0.193	0.117	0.250
Receiving 4+ food groups (at 6-23 months)	0.188	0.156	0.212
Child nutritional status			
Stunted (HAZ < -2)	0.681	0.678	0.683
Wasted (WAZ < -2)	0.062	0.079	0.049
Underweight (WHZ < -2)	0.339	0.395	0.298
Malnourished (MUAC $<$ 125 mm)	0.054	0.057	0.052
Child skills			
ASQ Communication score	37.61 (18.93)	41.02 (17.30)	35.16 (19.67)
ASQ Gross Motor score	35.03 (18.90)	34.37 (19.07)	35.51 (18.78)

Notes: The main sample for this table is all households surveyed at baseline. Each line corresponds to a variable. Columns (1) to (3) show mean (and standard deviation in parenthesis, if continuous) of the variable in the overall sample and in the two programme states separately. The variables are defined as follows: *Equivalised daily per capita expenditure* is obtained by summing 42 items of food and non-food expenditure covering all areas, equivalising the amount using the OECD scale, and converting it from Nigerian Naira to PPP US dollars at the 2014 rate. *Under poverty line* is a dummy for the above expenditure being below 1.9 USD PPP. *Woman and husband total monthly earnings* is obtained by summing the woman's and her husband's earnings across all work activities in the past 12 months, converting the amount to a monthly basis, and converting it to PPP US dollars as for expenditure. *Did not have enough food at some point in past year* is a dummy for households that report not having enough food to eat in the 4 seasons preceding the interview. *Work activity in past year* is a dummy indicating whether the woman has undertaken any paid or unpaid work activity in the past year (excluding housework and childcare for the household she resides in. *Cultivated land in past year* is a dummy for whether the woman has cultivated land at any point in the past year. *Hausa ethnicity* is a dummy for the household reporting Hausa ethnicity. *In polygamous marriage* is a dummy for the woman's marriage being polygamous. *Has a say on major HH purchases* is a dummy for whether the woman has any decision power on major household purchases (e.g. furniture). *Has a say on what food to buy* is a dummy for whether the woman has any decision power on what food is bought for household consumption. *Would go to [...]* are dummies derived from a list of people the woman would ask advice to about pregnancy or feeding a young child. *Put to the breast immediately* is a dummy for the child having been put to the breast in the first 30 minutes after birth. *Exclusively breastfed for at least 6m* is a dummy that takes value one only for children who have been exclusively breastfed for 6 months or more, and is not defined if the child is still being breastfed. *Appropriately breastfed* is a dummy indicating age-appropriate breastfeeding according to WHO guidelines (WHO, 2008), i.e. exclusive breastfeeding up to the age of 6 months and complementary breastfeeding from 6 to 23 months, and is not defined for children older that 23 months. *Receiving 4+ food groups* is a dummy indicating children who have received food from at least four among the following categories in the previous day: (i) Grains, roots and tubers; (ii) Legumes and nuts; (iii) Dairy products; (iv) Flesh foods; (v) Eggs; (vi) Vitamin-A rich fruits and veg.; (vii) Other fruits and veg; the indicator is not defined for children outside the 6-23 months age interval. *Stunted* is a dummy indicating children with height-for-age z-score (HAZ) under -2 SD. *Wasted* is a dummy indicating children with weight-for-age z-score (WAZ) under -2 SD. *Underweight* is a dummy indicating children with weight-for-height z-score (WHZ) under -2 SD. All z-scores are computed in accordance with WHO guidelines (WHO, 2009). *Malnourished* is a dummy indicating children with middle-upper-arm circumference (MUAC) lower than 125mm. *ASQ Communication score* and *ASQ Gross Motor score* are the raw scores from the communication and gross motor modules of the Ages and Stages Questionnaire, respectively.

Table 2.A2: Attrition

Dependent Variable Col. 1-4: Dummy = 1 if Woman Attrited
Dependent Variable Col. 5-7: Dummy = 1 if Old Child Attrited
Standard Errors Clustered by Village in Parentheses

	(1) Woman	(2) Woman	(3) Woman	(4) Woman (preg)	(5) Old child	(6) Old child	(7) Old child
T1	-0.0064	-0.0065	-0.0081	-0.0086	0.0037	0.0057	0.0270
	(0.0087)	(0.0088)	(0.0155)	(0.0172)	(0.0139)	(0.0136)	(0.0444)
T2	-0.0041	-0.0047	0.0077	0.0143	0.0235	0.0205	0.0288
	(0.0091)	(0.0092)	(0.0174)	(0.0196)	(0.0153)	(0.0153)	(0.0448)
Insecure village	0.9311***	0.9314***	0.9277***	0.9324***	0.8430***	0.8428***	0.8577***
	(0.0044)	(0.0045)	(0.0083)	(0.0091)	(0.0073)	(0.0083)	(0.0135)
Interactions							
T1 * insecure			0.0051	0.0046			-0.0143
			(0.0110)	(0.0117)			(0.0191)
T1 * household size			0.0002	0.0003			-0.0017
			(0.0017)	(0.0020)			(0.0032)
T1 * child age							-0.0002
							(0.0010)
T2 * insecure			0.0047	0.0018			-0.0405*
			(0.0120)	(0.0123)			(0.0198)
T2 * household size			-0.0017	-0.0023			-0.0003
			(0.0018)	(0.0021)			(0.0030)
T2 * child age							-0.0001
							(0.0010)
Baseline covariates							
Household size		-0.0026*	-0.0021	-0.0008		-0.0062**	-0.0056
		(0.0012)	(0.0015)	(0.0017)		(0.0020)	(0.0030)
Child age						-0.0026***	-0.0025***
						(0.0004)	(0.0007)
Rand. Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5433	5432	5432	3688	4162	4162	4162
Attrition rate	0.148	0.148	0.148	0.155	0.230	0.230	0.230
Joint p of covars		0.025	0.109	0.960		0.000	0.000
Joint p of interactions			0.828	0.762			0.429

Notes: The main sample for this table is all households surveyed at baseline. Each column shows estimates from a linear probability model where the dependent variable is attrition, and the independent variables are a varying set of treatment indicators, baseline covariates, and interactions. The main outcome variable, attrition, takes value 1 if the subject surveyed at baseline was not surveyed at midline. Columns (1) to (3) consider all women interviewed at baseline; column (4) restricts the sample to only women who were pregnant at baseline; columns (5) to (7) consider children aged 0-5 surveyed at baseline ("old" children). The number of nonmissing observations and the raw attrition rate are reported at the bottom of the table. The regressors used are defined as follows: T1 and T2 indicate households residing in communities receiving T1 and T2 treatments respectively; *Insecure village* denotes households residing in villages that could not be surveyed at midline because of security concerns for the enumerator teams; *Household size* is the number of people living in the household with common eating arrangements; *Child age* is the age of the "old" child at baseline, in months (included only in the columns regarding child attrition). Additional baseline covariates used in the regression model but omitted from the table are: number of children aged 0-2 in the household; number of children aged 3-5 in the household; Progress out of Poverty Index (PPI), a measure of household wealth; a dummy for polygamous marriage; and (for child attrition) a dummy for female child. All models also control for randomisation stratum, i.e. the tranche at which the PSU was randomised. Standard errors are clustered at the PSU level and displayed in parentheses below the point estimates. Significance: * (10%); ** (5%); *** (1%). At the bottom of the table, p-values are reported for F-tests of the joint significance of covariates and interaction terms, respectively.

Table 2.A3: Predictors of insecure villages – community level

Dependent Variable: Village Not Surveyed at Midline Due to Security Issues
Robust Standard Errors in Parentheses

	(1) Insecure village	(2) Insecure village
T1	0.0264	0.0246
	(0.0517)	(0.0502)
T2	-0.0307	-0.0393
	(0.0447)	(0.0436)
Baseline covariates		
Man-made shock in past 12m		0.1131*
		(0.0435)
Natural shock in past 12m		-0.1262
		(0.0683)
Other cash transfer		0.0723
		(0.0432)
Primary school		0.0002
		(0.0547)
Market		0.0060
		(0.0440)
Health facility		-0.0533
		(0.0431)
Mean PPI		-0.0023
		(0.0051)
Rand. Strata	Yes	Yes
Observations	208	208
Proportion insecure	0.087	0.087
Joint p of covars		0.043

Notes: The main sample for this table is all villages surveyed at baseline. Each column shows estimates from a linear probability model where the dependent variable is whether the villages could not be surveyed at midline because of security concerns for the enumerator teams; the independent variables are a varying set of treatment indicators and baseline covariates. The number of nonmissing observations and the rate of insecure villages are reported at the bottom of the table. The regressors used are defined as follows: T1 and T2 indicate villages receiving T1 and T2 treatments respectively; *Man-made shock in past 12m* indicates whether the village was exposed to curfews, violence, or widespread migration in the year before the listing survey; *Natural shock in past 12m* indicates whether the village was exposed to floods, droughts, or crop damage due to pests or disease in the year before the listing survey; *Other cash transfer* is a dummy for whether any other government or NGO cash transfer or cash benefit programme is concurrently operating in the village at the time of the listing survey; *Primary school*, *Market*, and *Health facility* are dummies for the presence of each in the village; *Mean PPI* is the mean level of the Progress out of Poverty Index (PPI), a measure of household wealth, among listed households in the village. All models also control for randomisation stratum, i.e. the tranche at which the PSU was randomised. Standard errors are clustered at the PSU level and displayed in parentheses below the point estimates. Significance: * (10%); ** (5%); *** (1%). At the bottom of the table, p-values are reported for F-tests of the joint significance of covariates.

Table 2.A4: Exposure to Low-Intensity Information Channels, Detail

Sample: Households with In Utero Pregnant Women at Baseline
Means, *p*-values in Brackets

	Women						Husbands						Women vs. Husbands		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	C	**T1**	**T2**	**pvalues of difference**			**C**	**T1**	**T2**	**pvalues of difference**			**pvalues of difference**		
	Mean	Mean	Mean	T1-C	T2-C	T2-T1	Mean	Mean	Mean	T1-C	T2-C	T2-T1	C	T1	T2
Low-intensity channels															
Posters	0.432	0.736	0.751	[.000]	[.000]	[.673]	0.376	0.603	0.623	[.000]	[.000]	[.641]	[.064]	[.000]	[.000]
Radio programme/ad	0.314	0.474	0.443	[.000]	[.002]	[.402]	0.547	0.681	0.630	[.004]	[.064]	[.228]	[.000]	[.000]	[.000]
Health talk	0.107	0.564	0.553	[.000]	[.000]	[.829]	0.113	0.243	0.242	[.000]	[.000]	[.988]	[.829]	[.000]	[.000]
Food demonstration	0.048	0.703	0.725	[.000]	[.000]	[.709]	0.007	0.060	0.051	[.000]	[.000]	[.647]	[.020]	[.000]	[.000]
Voice message		0.458	0.380			[.175]									
Randomisation strata				Yes	Yes	Yes				Yes	Yes	Yes	Yes	Yes	Yes
Observations	704	738	756				417	436	467						

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Each line corresponds to a variable. Columns (1) to (3) show mean of sampled women's exposure to BCC channels in each of the treatment groups. Columns (4) to (6) shows *p*-values of the hypothesis that the mean of the variable is equal across T1 and T2. These *p*-values are obtained by OLS regressions of the variable on dummies for the household residing in a T1 or T2 village, controlling for randomisation stratum (tranche) and clustering the standard errors at the village level. For each treatment arm, the number of observation is reported at the bottom of the table. This number is not the same for all variables in each group due to missingness and skip patterns. Columns (7) to (12) replicate the same analysis for husbands of the sampled women. Columns (13) to (15) report *p*-values for tests of equality between women's and husbands' exposure (when both are surveyed); these *p*-values are obtained by OLS regressions of each variables on dummies for treatment arm and respondent – woman vs. husband – and their interactions, controlling for randomisation stratum (tranche), and clustering the standard errors at the village level.

Table 2.A5: Takeup of Unconditional Cash Transfer, by State

Sample: Households with In Utero Pregnant Women at Baseline
Means, Standard Deviation in Parentheses, p-values in Brackets

	Jigawa				Zamfara			
	Control Mean (SD) (1)	**T1** Mean (SD) (2)	**T2** Mean (SD) (3)	**T1-T2 diff.** p-value (4)	**Control** Mean (SD) (5)	**T1** Mean (SD) (6)	**T2** Mean (SD) (7)	**T1-T2 diff.** p-value (8)
Cash payments takeup								
Ever received grant	0.054	0.952	0.954	[.972]	0.066	0.734	0.754	[.747]
If yes, still receiving grant at midline		0.902	0.866	[.474]		0.919	0.930	[.845]
Timing of first payment								
Age of New Child at first payment (months)		−0.31 (3.68)	0.20 (4.41)	[.166]		−0.33 (4.91)	0.04 (5.13)	[.994]
After birth		0.402	0.381	[.775]		0.315	0.326	[.813]
In same month of birth		0.149	0.144	[.734]		0.151	0.114	[.238]
During pregnancy		0.449	0.474	[.655]		0.534	0.557	[.509]
1st trimester		0.019	0.014	[.560]		0.044	0.023	[.338]
2nd trimester		0.136	0.107	[.245]		0.144	0.109	[.315]
3rd trimester		0.294	0.354	[.181]		0.346	0.425	[.059]
Intensity								
Number of payments		19.75 (2.92)	19.70 (3.63)	[.791]		19.87 (4.19)	19.60 (4.48)	[.624]
Frac. max payments (out of max. given child age)		0.82 (0.12)	0.84 (0.13)	[.175]		0.82 (0.14)	0.82 (0.14)	[.571]
Total grant amount received (USD PPP)		682.26 (102.09)	684.44 (132.46)	[.987]		693.56 (145.87)	679.08 (161.70)	[.712]
Randomisation strata				Yes				Yes
Observations	279	332	305		425	406	452	

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Each line corresponds to a variable. Columns (1) to (3) show mean (and standard deviation in parenthesis, if continuous) of the variable in each of the treatment groups; these are omitted for the control group apart from the overall take-up rate in the first row. Column (4) shows p-values of the hypothesis that the mean of the variable is equal across T1 and T2. These p-values are obtained by OLS regressions of the variable on dummies for the household residing in a T1 or T2 village, controlling for randomisation stratum (tranche) and clustering the standard errors at the village level. For each treatment arm, the number of observation is reported. This number is not the same for all variables in each group due to missingness and skip patterns. Columns (5)-(8) are repeated for Zamfara. The variables are defined as follows: *Ever received grant* is a dummy for the woman reporting having aver participated in the CDGP cash transfer programme. *Still receiving grant at midline* is a dummy defined only for women who have ever received the grant, taking value 1 if they are still receiving the grant at the midline interview. *Age of New Child at first payment* is the age of the surveyed child born after the baseline survey at the time the first payment was received by the mother; it's computed by comparing the reported month of birth of the child (from the household questionnaire) with the reported month when the mother received her first CDGP payment (from the implementation dataset); negative ages imply payments received before birth. *During pregnancy*, *1st trimester*, ..., *Infeasible* are dummies indicating children who (according to the above age computation) have received the payment during pregnancy, in each of the 3 trimesters, in the same months as they were born, after they had been born, and for which the age at first payment takes an infeasible value (below -9). *Number of payments* is the number of payments the mother received by the midline interview, according to the implementation dataset. *Frac. max payments* is the ratio between the number of payments received and the maximum number of payments to which the mother is entitled to; the denominator is computed by subtracting the month of first payment from the month when the child will turn 2 years old and the cash grant ends. *Total grant amount received* is the money amount of cash received up to the midline interview; this is calculated using data from the implementation dataset; it is then deflated to August 2014 using the monthly rural CPI index published by the Central Bank of Nigeria (Central Bank of Nigeria, 2017), and finally converted using the PPP US dollar rate for August 2014.

Table 2.A6: Exposure to Low- and High-Intensity Information, by State

Sample: Households with In Utero Pregnant Women at Baseline
Means, p-values in Brackets

	Women						Husbands						Women vs. Husbands		
	C	T1	T2	pvalues of difference			C	T1	T2	pvalues of difference			pvalues of difference		
	Mean	Mean	Mean	T1-C	T2-C	T2-T1	Mean	Mean	Mean	T1-C	T2-C	T2-T1	C	T1	T2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Jigawa – Low-intensity channels															
None	0.333	0.033	0.026	[.000]	[.000]	[.691]	0.310	0.120	0.134	[.000]	[.001]	[.720]	[.510]	[.001]	[.000]
At least one	0.667	0.967	0.974	[.000]	[.000]	[.691]	0.690	0.880	0.866	[.000]	[.001]	[.720]	[.510]	[.001]	[.000]
All	0.007	0.244	0.252	[.000]	[.000]	[.884]	0.006	0.047	0.043	[.013]	[.020]	[.910]	[.258]	[.000]	[.000]
Panel B: Jigawa – High-intensity channels															
None	0.953	0.268	0.177	[.000]	[.000]	[.019]	0.981	0.807	0.829	[.000]	[.000]	[.663]	[.364]	[.000]	[.000]
All	0.007	0.166	0.223	[.000]	[.000]	[.100]	0.000	0.057	0.032	[.013]	[.012]	[.340]	[.342]	[.000]	[.000]
Support group		0.696	0.813			[.005]		0.156	0.096			[.147]		[.000]	[.000]
Says 1:1 counselling available in village		0.599	0.620			[.718]		0.479	0.449			[.648]		[.002]	[.001]
If yes, tried to obtain 1:1 counselling		0.407	0.386			[.604]		0.228	0.250			[.556]		[.003]	[.034]
If yes, obtained 1:1 counselling		0.827	0.973			[.004]		0.857	0.952			[.270]		[.517]	[.888]
Panel C: Zamfara – Low-intensity channels															
None	0.456	0.145	0.111	[.000]	[.000]	[.338]	0.363	0.176	0.239	[.000]	[.022]	[.129]	[.022]	[.386]	[.000]
At least one	0.544	0.855	0.889	[.000]	[.000]	[.338]	0.637	0.824	0.761	[.000]	[.022]	[.129]	[.022]	[.386]	[.000]
All	0.005	0.091	0.044	[.000]	[.007]	[.031]	0.004	0.029	0.029	[.013]	[.045]	[.958]	[.780]	[.000]	[.315]
Panel D: Zamfara – High-intensity channels															
None	0.922	0.611	0.437	[.000]	[.000]	[.010]	0.920	0.885	0.879	[.165]	[.144]	[.751]	[.997]	[.000]	[.000]
All	0.007	0.059	0.102	[.010]	[.000]	[.146]	0.008	0.004	0.029	[.559]	[.061]	[.022]	[.983]	[.009]	[.000]
Support group		0.377	0.548			[.010]		0.094	0.096			[.883]		[.000]	[.000]
Says 1:1 counselling available in village		0.190	0.333			[.026]		0.152	0.296			[.009]		[.316]	[.241]
If yes, tried to obtain 1:1 counselling		0.403	0.360			[.906]		0.189	0.205			[.875]		[.123]	[.019]
If yes, obtained 1:1 counselling		0.935	0.981			[.369]		0.857	0.882			[.824]		[.622]	[.188]
Randomisation strata				Yes	Yes	Yes				Yes	Yes	Yes	Yes	Yes	Yes
Observations	704	738	756				417	436	467						

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). Each line corresponds to a variable. Columns (1) to (3) show mean of sampled women's exposure to BCC channels in each of the treatment groups. Columns (4) to (6) shows p-values of the hypothesis that the mean of the variable is equal across T1 and T2. These p-values are obtained by OLS regressions of the variable on dummies for the household residing in a T1 or T2 village, controlling for randomisation stratum (tranche) and clustering the standard errors at the village level. For each treatment arm, the number of observation is reported at the bottom of the table. This number is not the same for all variables in each group due to missingness and skip patterns. Columns (7) to (12) replicate the same analysis for husbands of the sampled women. Columns (13) to (15) report p-values for tests of equality between women's and husbands' exposure (when both are surveyed); these p-values are obtained by OLS regressions of each variables on dummies for treatment arm and respondent – woman vs. husband – and their interactions, controlling for randomisation stratum (tranche), and clustering the standard errors at the village level. Low-intensity channels are posters, radio programmes/ads, health talks, food demonstrations, and voice messages. Men are not surveyed about receiving voice messages, so their maximum number of low-intensity channels is four – while it's 5 for women. High-intensity channels are support groups and one-to-one counselling. *None* of the low-intensity channels is a dummy for the respondent reporting not having been exposed to any of the channels. *At least one* is a dummy for the respondent reporting having been exposed to at least one of the five low-intensity channels (four for men). *All* is a dummy for being exposed to all of the channels. Participation to one-to-one counselling is surveyed sequentially: first, respondents are asked if such activity is available in their village; in case of a positive answer, they are asked whether they ever tried to access it; in case of a positive answer, they are asked if they were able to obtain it.

Table 2.A7: First Stage for Control Function

Dependent Variable: Age of New Child at Midline
Sample: In Utero New Child
Standard Errors Clustered by Village in Parentheses

	(1) Month of Int.	(2) Week of Int.	(3) Day of Int.	(4) Day of Int. (Males)	(5) Day of Int. (Females)
Dependent var.: age of new child (months)					
Treated	-0.740***	-0.511***	-0.612***	-0.661***	-0.536**
	(0.129)	(0.124)	(0.123)	(0.162)	(0.162)
November	0.562***				
	(0.121)				
Week 3		0.761***			
		(0.149)			
Week 4		0.779***			
		(0.156)			
Week 5		0.953***			
		(0.165)			
Week 6		0.822***			
		(0.158)			
Week 7-8		1.926***			
		(0.256)			
Day of interview			0.034***	0.033***	0.036***
			(0.004)	(0.006)	(0.006)
Randomisation Strata	✓	✓	✓	✓	✓
Region effects	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓
Observations	2155	2155	2155	1120	1035
F-stat for instrument	21.43	16.71	60.61	32.54	33.92
Adj R2 difference from instrument	.01	.033	.026	.025	.028

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The dependent variable is the age of the new child in months. *November* is a dummy for whether the midline interview took place in November 2016 (reference: October 2016). *Week 3* to *Week 7-8* are dummies for each week of interview (reference: Week 1-2, October 3rd to 16th 2016). *Day of interview* is a continuous indicator of day of interview. Standard errors clustered at the village level are reported in parentheses. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline, and a dummy for the specialist who measured the child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A8: New child anthropometrics – Control function estimates

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

		Raw	Age-adjusted	Age-adjusted + CF		Age-adj. + interaction	Age-adj. + interaction + CF	
	C Mean	Effect	Effect	Effect	p of CF polyn	Effect	Effect	p of CF polyn
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Continuous indicators								
Summary index (using †)	−0.00 (1.00)	0.12** (0.06)	0.08 (0.06)	0.08 (0.06)	0.623	0.07 (0.06)	0.07 (0.06)	0.614
Height-for-age (HAZ)†	−2.72 (1.12)	0.21*** (0.06)	0.12** (0.06)	0.16** (0.06)	0.024	0.11* (0.06)	0.16** (0.07)	0.023
Weight-for-age (WAZ)†	−1.86 (1.08)	0.05 (0.06)	0.02 (0.06)	0.03 (0.06)	0.022	0.01 (0.06)	0.02 (0.06)	0.019
Weight-for-height (WHZ)	−0.64 (1.05)	−0.10* (0.06)	−0.07 (0.06)	−0.09 (0.06)	0.029	−0.08 (0.06)	−0.09 (0.06)	0.021
Middle Upp. Arm Circumf. (MUAC)†	136.47 (11.77)	−0.06 (0.65)	0.12 (0.65)	−0.32 (0.68)	0.202	−0.06 (0.64)	−0.50 (0.68)	0.204
Thresholds								
Summary index (using †)	−0.00 (1.00)	−0.05 (0.06)	−0.01 (0.06)	−0.03 (0.06)	0.170	−0.00 (0.06)	−0.02 (0.06)	0.217
Stunted (HAZ<-2)†	73.47	−4.82** (2.45)	−1.76 (2.43)	−3.34 (2.56)	0.015	−1.45 (2.48)	−3.05 (2.64)	0.020
Severely Stunted (HAZ<-3)	39.37	−4.96** (2.42)	−2.40 (2.34)	−4.13 (2.69)	0.265	−1.93 (2.30)	−3.62 (2.63)	0.267
Underweight (WAZ<-2)†	42.51	−0.23 (2.61)	1.17 (2.62)	1.02 (2.99)	0.096	1.51 (2.58)	1.33 (2.99)	0.087
Wasted (WHZ<-2)	9.99	3.42** (1.49)	2.76* (1.53)	2.27 (1.71)	0.063	3.13** (1.54)	2.63 (1.71)	0.074
Malnourished (MUAC<125mm)†	14.00	0.22 (1.78)	−0.07 (1.79)	−0.19 (1.89)	0.983	0.34 (1.76)	0.21 (1.86)	0.991
Age adjustment			✓	✓		✓	✓	
Age-treatment interaction						✓	✓	
Control function for age				✓			✓	

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Height-for-age, weight-for-age, and weight-for-height indicators are computed with reference to the WHO growth standards (WHO, 2009). Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Col. (4) shows the same as col. (3), but the estimates are also corrected for the possible endogeneity of the age of the child with respect to treatment. The correction is implemented by control functions. In a first stage, the age of the child is regressed on all covariates and date of midline interview (the exogenous instrument); residuals from the first stage are then included in the regression for the outcome, in a flexible cubic polynomial form. Standard errors are computed by bootstrap with 1,000 repetitions. Col. (5) reports the p-value of the test that the coefficients on the control function terms are equal to zero. Col. (6) shows the same as col. (2), but age is interacted with treatment status. The reported estimate is the ITT averaged across age groups, with weights proportional to the sample size in each age group. Col. (7) shows the same as col. (6), but the estimates are also corrected for endogeneity in the same way as col. (4). Col. (5) reports the p-value of the test that the coefficients on the control function terms are equal to zero. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A9: New child anthropometrics – Unstandardised and internally standardised outcomes

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	Raw Effect	Age-adjusted effect
	(1)	(2)	(3)
Height (cm)	76.57 (3.60)	0.01 (0.17)	0.31* (0.18)
Height (internal Z)	−0.02 (1.00)	0.09* (0.05)	0.10** (0.05)
Weight (kg)	9.29 (1.27)	−0.07 (0.07)	0.01 (0.07)
Weight (internal Z)	−0.01 (0.99)	0.02 (0.05)	0.02 (0.05)
Middle upper arm circumference (mm)	136.47 (11.77)	−0.06 (0.65)	0.12 (0.65)
Middle upper arm circumference (internal Z)	−0.00 (1.00)	0.01 (0.06)	0.01 (0.06)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Height-for-age, weight-for-age, and weight-for-height indicators are computed with reference to the WHO growth standards (WHO, 2009). Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A10: New child height – Robustness to age groupings (no control functions)

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

				Age groupings				
	Linear age	Quadr. age	Cubic age	2 groups	3 groups	4 groups	5 groups	Month dummies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Height-for-age (HAZ)	0.11^{*}	0.11^{*}	0.10^{*}	0.15^{***}	0.14^{**}	0.12^{**}	0.12^{**}	0.10^{*}
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Height (cm)	0.38^{**}	0.37^{**}	0.36^{**}	0.23	0.26	0.31^{*}	0.31^{*}	0.36^{**}
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.18)	(0.17)	(0.17)
Height (internal Z)	0.11^{**}	0.11^{**}	0.11^{**}	0.10^{**}	0.10^{**}	0.10^{**}	0.10^{**}	0.11^{**}
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). We also exclude children aged 14 and 27 months, since the sample size in those age-treatment cells is small. The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Height-for-age is computed with reference to the WHO growth standards (WHO, 2009). Internal standardisations are obtained by standardising height in centimeters, weight in kilograms, or MUAC in mm, at each month of age using mean and variance profiles smoothed by kernel-weighted local means. Each column shows ITT effects adjusted for the children's age. Col. (1) adjusts for age in months linearly. Col. (2) and (3) use a quadratic and cubic in age in months, respectively. Col. (4) uses two age groups: 15-21, 22-26. Col. (5) uses three age groups: 15-20, 21-22, 23-26. Col. (6) uses four age groups: 15-17, 18-20, 21-23, 24-26. Col. (7) uses five age groups: 15-18, 19-20, 21-22, 23, 24-26. Col. (8) uses month of age dummies between 15 and 26 months. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A11: New child height – Robustness to age groupings (with control functions)

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

	2 age groups		3 age groups		4 age groups		5 age groups		Month dummies	
	Effect	p of CF polyn.	Effect	p of CF polyn.	Effect	p of CF polyn.	Effect	p of CF polyn.	Effect	p of CF polyn.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Height-for-age (HAZ)	0.20*** (0.06)	0.000	0.18*** (0.06)	0.000	0.16** (0.07)	0.023	0.13** (0.07)	0.049	0.04 (0.08)	0.047
Height (cm)	0.03 (0.18)	0.000	−0.00 (0.18)	0.000	−0.03 (0.19)	0.000	−0.06 (0.19)	0.000	0.11 (0.24)	0.047
Height (internal Z)	0.09* (0.05)	0.173	0.07 (0.05)	0.080	0.06 (0.05)	0.036	0.04 (0.06)	0.028	0.04 (0.07)	0.023

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). We also exclude children aged 14 and 27 months, since the sample size in those age-treatment cells is small. The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Height-for-ageis computed with reference to the WHO growth standards (WHO, 2009). Internal standardisations are obtained by standardising height in centimeters, weight in kilograms, or MUAC in mm, at each month of age using mean and variance profiles smoothed by kernel-weighted local means. Each pair of columns shows in the first column ITT effects adjusted for the children's age groups, which are also interacted with treatment status. The reported estimate is the ITT averaged across age groups, with weights proportional to the sample size in each age group. The second column of each pair shows the p-value of the test that the coefficients on the control function terms are equal to zero. Col. (1)-(2) use two age groups: 15-21, 22-26. Col. (3)-(4) use three age groups: 15-20, 21-22, 23-26. Col. (5)-(6) use four age groups: 15-17, 18-20, 21-23, 24-26. Col. (7)-(8) use five age groups: 15-18, 19-20, 21-22, 23, 24-26. Col. (9)-(10) uses groups corresponding to each month of age between 15 and 26 months. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A12: Old child anthropometrics – Unstandardised outcomes

Sample: Households with In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	Raw Effect	Age-adjusted effect
	(1)	(2)	(3)
Height (cm)	113.87 (108.98)	−3.93 (6.28)	−4.01 (6.27)
Weight (kg)	29.80 (119.17)	−4.55 (6.82)	−4.73 (6.83)
Middle upper arm circumference (mm)	162.60 (95.47)	−1.90 (5.68)	−2.13 (5.69)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline, the new child is estimated to have been in utero at baseline, and an old child was surveyed at baseline ($N = 1,620$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Internal standardisations are obtained by standardising height in centimeters, weight in kilograms, or MUAC in mm, at each month of age using mean and variance profiles smoothed by kernel-weighted local means. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for a quadratic polynomial of the age of the child at midline. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A13: New child communication and motor skills – Control function estimates

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

		Raw	**Age-adjusted**	**Age-adjusted + CF**		**Age-adj. + interaction**	**Age-adj. + interaction + CF**	
	C Mean	Effect	Effect	Effect	p of CF polyn	Effect	Effect	p of CF polyn
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Communication skills (Z)	−0.00 (1.00)	0.08 (0.06)	0.11* (0.06)	0.05 (0.07)	0.001	0.10* (0.06)	0.05 (0.07)	0.001
Communication skills in normal range	30.76	5.74** (2.36)	5.73** (2.38)	5.30** (2.64)	0.002	5.14** (2.33)	4.86* (2.62)	0.002
Gross motor skills (Z)	0.00 (1.00)	0.09 (0.06)	0.07 (0.06)	0.04 (0.07)	0.487	0.06 (0.06)	0.03 (0.07)	0.478
Gross motor skills in normal range	40.73	4.14 (3.01)	2.91 (3.06)	3.44 (3.47)	0.961	2.83 (3.07)	3.37 (3.49)	0.956
Age adjustment			✓	✓		✓	✓	
Age-treatment interaction						✓	✓	
Control function for age				✓			✓	

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Communication and motor skills Z scores are obtained by standardising raw scores using the mean and standard deviation in the control group at midline. 'Normal ranges' are binary indicators that take value 1 if the score falls in the normal category based on the reference population. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Col. (4) shows the same as col. (3), but the estimates are also corrected for the possible endogeneity of the age of the child with respect to treatment. The correction is implemented by control functions. In a first stage, the age of the child is regressed on all covariates and date of midline interview (the exogenous instrument); residuals from the first stage are then included in the regression for the outcome, in a flexible cubic polynomial form. Standard errors are computed by bootstrap with 1,000 repetitions. Col. (5) reports the p-value of the test that the coefficients on the control function terms are equal to zero. Col. (6) shows the same as col. (2), but age is interacted with treatment status. The reported estimate is the ITT averaged across age groups, with weights proportional to the sample size in each age group. Col. (7) shows the same as col. (6), but the estimates are also corrected for endogeneity in the same way as col. (4). Col. (5) reports the p-value of the test that the coefficients on the control function terms are equal to zero. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A14: New child vaccinations

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean	Effect
	(1)	(2)
BCG	24.72	12.54*** (2.84)
Polio	93.54	1.25 (1.49)
Polio at birth	43.68	5.01* (2.81)
3+ polio shots	89.20	0.33 (1.82)
DPT	14.61	4.33** (2.17)
3+ DPT shots	1.72	0.66 (0.76)
Measles	35.67	13.38*** (3.22)
Hepatitis B	10.25	6.96*** (1.98)
Yellow fever	17.84	14.10*** (2.72)
All basic vaccinations	0.70	1.10* (0.58)
None of the basic vaccinations	5.90	−1.05 (1.43)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Basic vaccinations are defined as BCG, measles, and 3 doses each of DPT and polio vaccine (excluding polio vaccine given at birth). Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.A15: Antenatal Care, Details

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean (1)	Effect (2)
***Panel A**: Antenatal care for New Child*		
Received antenatal care†	60.96	8.96** (3.86)
Times received antenatal care	2.79 (2.94)	0.64*** (0.24)
Reason for not seeking antenatal care for New Child		
Travel cost too high	4.07	−0.79 (1.17)
No permission	5.90	−3.01*** (1.10)
Saw no reason to seek it	29.63	−7.41** (3.24)
Other reason	2.81	0.09 (0.82)
***Panel B**: Antenatal care for current pregnancy*		
Received antenatal care	16.30	17.41*** (3.44)
Months pregnant when first got AC	3.47 (1.75)	0.00 (0.24)
Transport cost	199.18 (266.18)	66.75 (49.38)
Treatment cost	440.95 (680.55)	−32.88 (123.48)
Received iron supplements	77.78	12.38* (7.06)
Received folic acid	84.44	−4.78 (7.51)
Received tetanus shot	62.22	6.20 (7.25)
Received deworming drugs	36.11	2.80 (8.43)
Received malaria drugs	68.89	4.13 (8.51)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ANCOVA ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.A16: Effects on village-level availability of health services
Sample: Evaluation Villages
ITT Estimates, Robust Standard Errors in Parentheses

	C Mean (1)	**Effect** (2)
Distance from health facility		
Available in village	33.87	11.71 (7.43)
Is more than 1h away on foot	38.71	−11.07 (6.98)
Euclidean distance (GPS, km)	1.44 (1.53)	0.13 (0.24)
Availability of services		
Antenatal care	83.61	5.63 (5.21)
Postnatal care	84.75	−5.24 (5.90)
Delivery of babies	75.00	−5.30 (6.42)
Immunization	96.72	0.61 (2.69)
Healthy diet counselling	73.21	13.51** (5.90)
Availability of staff		
Doctor	35.09	6.50 (7.18)
Nurse/Midwife	54.24	1.55 (7.57)
CHEW	96.61	0.53 (2.68)

Notes: The main sample for this table is all villages surveyed at midline ($N = 192$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline. Col. (2) shows the ITT effect. Heteroskedasticity-robust standard errors are in parentheses. Means and effects on binary variables are reported in percentage points. All estimates control for region effects and randomisation tranche. Significance: * (10%); ** (5%); *** (1%).

Table 2.A17: Borrowing and saving

Sample: Households with In Utero Pregnant Women at Baseline
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean (1)	Effect (2)
Borrowing		
Any household member borrowing money	24.22	−1.91 (2.48)
from bank, savings association, microfinance, or NGO	2.14	0.20 (0.89)
from family members or friends	20.90	−3.67 (2.83)
from shop, landlord, moneylender, or other	9.40	2.06 (2.38)
Tot. value of borrowing (PPP USD)	32.60 (127.42)	−15.61 (9.74)
Saving		
Any household member saving money (incl. in-kind)	64.07	2.76 (2.26)
at bank, savings association, microfinance, or NGO	9.93	0.06 (1.39)
at home	35.81	1.70 (3.01)
at informal savings group	8.54	2.80 (2.01)
Tot. value of savings (incl. in-kind, PPP USD)	244.90 (631.49)	−25.07 (67.81)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ANCOVA ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at baseline. Col. (2) shows the ANCOVA ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the new child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A18: Home production

Sample: Households with In Utero Pregnant Women at Baseline
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	C Mean (1)	Effect (2)
Crops cultivated by woman or husband in past 12 months		
Grains, tubers, roots	92.98	−0.52 (1.73)
Dark green leafy vegetables	0.00	0.27 (0.23)
Other fruit and vegetables	2.67	−1.60 (1.41)
Nuts, beans, and seeds	35.67	−2.86 (2.66)
Animals owned by household at interview		
Any milk-producing animal (female cow, goat, or sheep)	70.65	−0.48 (2.02)
Number of milk-producing animals (female cow, goat, or sheep)	3.94 (5.58)	−0.11 (0.41)
Any commonly eaten animal (cow/bull, calf, goat, sheep)	75.56	1.93 (1.80)
Number of commonly eaten animals (cow/bull, calf, goat, sheep)	5.77 (8.44)	−0.05 (0.69)

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ANCOVA ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at baseline. Col. (2) shows the ANCOVA ITT effect. Standard errors are in parentheses (clustered at the village level). Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the new child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A19: Use of cash grant

Sample: Households with In Utero Pregnant Women at Baseline
Means, p-values

	Woman			Husband		
	T1 Mean	T2 Mean	p of diff. T1-T2	T1 Mean	T2 Mean	p of diff. T1-T2
	(1)	(2)	(3)	(4)	(5)	(6)
What used most of the grant for						
Buying food for the household	63.84	64.40	0.993	66.86	61.95	0.154
Buying food for children	25.73	23.89	0.478	21.43	24.16	0.480
Other	10.42	11.71	0.265	11.71	13.88	0.303
What else used the grant for						
Buying food for children	26.06	26.11	0.822	27.43	25.71	0.601
Nothing else	24.76	25.79	0.581	24.00	21.59	0.652
Buying food for the household	18.89	17.25	0.817	14.86	15.42	0.671
Savings	16.78	15.35	0.429	10.00	9.51	0.664
Health expenses for children	9.77	10.76	0.356	10.00	8.48	0.937
Other	29.80	26.42	0.362	21.71	15.42	0.113

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Columns (1) and (2) show the mean of the indicator at midline in the T1 and T2 groups, respectively, as reported by the woman. Column (3) reports the p-value for the hypothesis that the mean is the same in T1 and T2. Columns (4) to (6) report the same statistics, but based on the husband's report. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the new child at midline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A20: Mediation Analysis for Continuous Outcomes

Sample: Households with In Utero Pregnant Women at Baseline
Gelbach Decomposition Estimates, Standard Errors Clustered by Village in Parentheses

	Raw		Age-adjusted	
	Coefficient	Fraction mediated	Coefficient	Fraction mediated
	(1)	(2)	(3)	(4)
Outcome: ***Height-for-age (HAZ)***				
Unconditional effect (ITT)	0.1955*** (0.0611)		0.1011* (0.0584)	
Conditional effect	0.1133* (0.0646)		0.0308 (0.0620)	
Mediated by:				
Mother's knowledge	0.0547** (0.0264)	28.0	0.0579** (0.0252)	57.2
Breastfeeding practices	−0.0022 (0.0202)	−1.1	−0.0133 (0.0188)	−13.1
Antenatal care practices	0.0171* (0.0088)	8.7	0.0112 (0.0069)	11.1
Child dietary diversity	0.0092 (0.0081)	4.7	0.0113 (0.0080)	11.2
Household expenditure	0.0016 (0.0024)	0.8	0.0014 (0.0021)	1.4
Mother's work	0.0018 (0.0046)	0.9	0.0018 (0.0043)	1.7
Total mediated	0.0822*** (0.0284)	42.1	0.0703*** (0.0272)	69.5
Outcome: ***Communication skills Z***				
Unconditional effect (ITT)	0.0702 (0.0596)		0.1022* (0.0598)	
Conditional effect	0.0082 (0.0640)		0.0384 (0.0640)	
Mediated by:				
Mother's knowledge	0.0262 (0.0248)	37.3	0.0237 (0.0243)	23.2
Breastfeeding practices	−0.0017 (0.0208)	−2.5	0.0046 (0.0202)	4.5
Antenatal care practices	0.0221** (0.0111)	31.5	0.0208* (0.0111)	20.3
Child dietary diversity	0.0208*** (0.0074)	29.6	0.0195*** (0.0073)	19.1
Household expenditure	0.0004 (0.0018)	0.5	0.0007 (0.0017)	0.6
Mother's work	−0.0057 (0.0046)	−8.1	−0.0055 (0.0043)	−5.4
Total mediated	0.0620** (0.0297)	88.3	0.0638** (0.0295)	62.4

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports decompositions of the ITT effect according to the methodology in Gelbach (2016). For each outcome, Col. (1) shows the unconditional effect (corresponding to the ITT) derived from the restricted model, i.e. the one estimated without mediators; the conditional effect, which is the equivalent of the ITT in the full model, i.e. with the set of mediators added as explanatory variables; and the decomposition of the difference between unconditional and conditional effect explained by each group of mediators. Col. (2) shows the fraction of the difference explained by all mediators and by each mediator group. Standard errors are in parentheses (clustered at the village level). The mediator groups are defined as follows: *Mother's knowledge* [Index of maternal knowledge about ECD practices, see Table 2.9] *Breastfeeding practices* related to the new child [Fed colostrum in the first hour, put to the breast immediately, exclusively breastfed for 6-7 months], *Antenatal care practices* related to the new child [Received antenatal care, born at health facility], *Child dietary diversity* [Dietary diversity index, see Table 2.11], *Total HH expenditure*, *Mother's work* [Any work activity in past 12 months, number of activities]. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, and the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A21: Mediation Analysis for Threshold Outcomes

Sample: Households with In Utero Pregnant Women at Baseline
Gelbach Decomposition Estimates, Standard Errors Clustered by Village in Parentheses

	Raw		Age-adjusted	
	Coefficient	Fraction mediated	Coefficient	Fraction mediated
	(1)	(2)	(3)	(4)
Outcome: ***Stunted (WAZ*** < -2***)***				
Unconditional effect (ITT)	-0.0497^{**} (0.0247)		-0.0174 (0.0247)	
Conditional effect	-0.0293 (0.0266)		-0.0007 (0.0266)	
Mediated by:				
Mother's knowledge	-0.0098 (0.0108)	19.8	-0.0109 (0.0104)	62.6
Breastfeeding practices	-0.0070 (0.0091)	14.1	-0.0033 (0.0085)	19.0
Antenatal care practices	-0.0034 (0.0026)	6.9	-0.0017 (0.0021)	9.8
Child dietary diversity	-0.0003 (0.0030)	0.6	-0.0009 (0.0030)	5.2
Household expenditure	-0.0000 (0.0009)	0.1	-0.0000 (0.0008)	0.3
Mother's work	0.0002 (0.0021)	-0.4	0.0002 (0.0020)	-1.1
Total mediated	-0.0204^{**} (0.0102)	41.0	-0.0166^{*} (0.0099)	95.8
Outcome: ***Communication skills referral/monitoring***				
Unconditional effect (ITT)	-0.0541^{**} (0.0242)		-0.0541^{**} (0.0244)	
Conditional effect	-0.0278 (0.0262)		-0.0289 (0.0263)	
Mediated by:				
Mother's knowledge	-0.0083 (0.0116)	15.4	-0.0081 (0.0114)	14.9
Breastfeeding practices	-0.0001 (0.0103)	0.1	-0.0006 (0.0100)	1.1
Antenatal care practices	-0.0086^{**} (0.0042)	15.9	-0.0077^{*} (0.0041)	14.2
Child dietary diversity	-0.0078^{***} (0.0029)	14.3	-0.0076^{***} (0.0029)	14.0
Household expenditure	-0.0010 (0.0011)	1.8	-0.0009 (0.0011)	1.7
Mother's work	-0.0006 (0.0020)	1.1	-0.0004 (0.0019)	0.8
Total mediated	-0.0264^{**} (0.0134)	48.7	-0.0253^{*} (0.0132)	46.7

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports decompositions of the ITT effect according to the methodology in Gelbach (2016). For each outcome, Col. (1) shows the unconditional effect (corresponding to the ITT) derived from the restricted model, i.e. the one estimated without mediators; the conditional effect, which is the equivalent of the ITT in the full model, i.e. with the set of mediators added as explanatory variables; and the decomposition of the difference between unconditional and conditional effect explained by each group of mediators. Col. (2) shows the fraction of the difference explained by all mediators and by each mediator group. Standard errors are in parentheses (clustered at the village level). The mediator groups are defined as follows: *Mother's knowledge* [Index of maternal knowledge about ECD practices, see Table 2.9] *Breastfeeding practices* related to the new child [Fed colostrum in the first hour, put to the breast immediately, exclusively breastfed for 6-7 months], *Antenatal care practices* related to the new child [Received antenatal care, born at health facility], *Child dietary diversity* [Dietary diversity index, see Table 2.11], *Total HH expenditure*, *Mother's work* [Any work activity in past 12 months, number of activities]. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, and the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

2.9.4 Results by Child Gender

Table 2.A22: New child communication and motor skills, by gender

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	Males				Females				M-F Diff.	
	C Mean	Effect	Effect	Effect	C Mean	Effect	Effect	Effect	p	p
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Communication skills (Z)	−0.07 (0.98)	0.06 (0.07)	0.08 (0.07)	0.01 (0.08)	0.08 (1.02)	0.10 (0.08)	0.13* (0.07)	0.08 (0.08)	0.530	0.566
Communication skills in normal range	28.24	5.26* (2.96)	5.04* (3.01)	4.14 (3.34)	33.74	5.84* (3.26)	5.99* (3.24)	5.97* (3.45)	0.859	0.903
Gross motor skills (Z)	0.01 (1.03)	0.09 (0.07)	0.08 (0.08)	0.06 (0.08)	−0.01 (0.96)	0.07 (0.08)	0.05 (0.08)	0.01 (0.09)	0.944	0.986
Gross motor skills in normal range	43.26	2.83 (3.67)	1.98 (3.72)	2.67 (4.20)	37.73	5.39 (3.93)	3.73 (4.02)	3.85 (4.61)	0.439	0.460
Age adjustment			✓	✓			✓	✓		✓
Control function				✓				✓		

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for male children. Col. (2) shows the ITT effect for males. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Col. (4) shows the same as col. (3), but the estimates are also corrected for the possible endogeneity of the age of the child with respect to treatment. The correction is implemented by control functions. In a first stage, the age of the child is regressed on all covariates and date of midline interview (the exogenous instrument); residuals from the first stage are then included in the regression for the outcome, in a flexible cubic polynomial form. Standard errors are computed by bootstrap with 1,000 repetitions. Col. (5)-(8) repeat for female children. Col. (10)-(11) show p-values of the hypothesis that the effect on females and males is the same, with and without age adjustment. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A23: New child anthropometrics, by gender

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

	Males				**Females**				**M-F Diff.**	
	C Mean (1)	Effect (2)	Effect (3)	Effect (4)	C Mean (5)	Effect (6)	Effect (7)	Effect (8)	p (9)	p (10)
Continuous indicators										
Height-for-age (HAZ)	-2.85 (1.19)	0.24*** (0.09)	0.14* (0.08)	0.18* (0.09)	-2.57 (1.03)	0.15** (0.07)	0.08 (0.06)	0.12* (0.07)	0.429	0.377
Weight-for-age (WAZ)	-1.91 (1.05)	0.07 (0.07)	0.02 (0.07)	0.04 (0.08)	-1.80 (1.11)	0.01 (0.08)	-0.01 (0.08)	0.01 (0.09)	0.645	0.626
Middle Upp. Arm Circumf. (MUAC)	137.67 (11.50)	0.00 (0.82)	0.07 (0.82)	-0.47 (0.87)	135.06 (11.95)	-0.35 (0.85)	-0.01 (0.84)	-0.40 (0.92)	0.691	0.697
Thresholds										
Stunted (HAZ<-2)	74.80	-4.44 (3.17)	-1.34 (3.06)	-3.25 (3.51)	71.91	-5.06 (3.27)	-2.24 (3.27)	-3.78 (3.39)	0.864	0.891
Underweight (WAZ<-2)	42.71	1.50 (3.45)	3.10 (3.53)	3.53 (3.88)	42.28	-1.52 (3.37)	-0.59 (3.36)	-1.31 (3.87)	0.432	0.449
Malnourished (MUAC<125mm)	12.04	0.30 (2.09)	0.38 (2.08)	0.71 (2.22)	16.31	0.43 (2.61)	-0.26 (2.69)	-1.01 (3.05)	0.849	0.860
Age adjustment			✓	✓			✓	✓		✓
Control function				✓				✓		

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for male children. Col. (2) shows the ITT effect for males. Standard errors are in parentheses (clustered at the village level). Col. (3) shows the same as col. (2), but the estimates are adjusted for the age of the child at midline. The adjustment is made using dummies for four different age ranges in months: 14-17, 18-20, 21-23, 24-27. Col. (4) shows the same as col. (3), but the estimates are also corrected for the possible endogeneity of the age of the child with respect to treatment. The correction is implemented by control functions. In a first stage, the age of the child is regressed on all covariates and date of midline interview (the exogenous instrument); residuals from the first stage are then included in the regression for the outcome, in a flexible cubic polynomial form. Standard errors are computed by bootstrap with 1,000 repetitions. Col. (5)-(8) repeat for female children. Col. (10)-(11) show p-values of the hypothesis that the effect on females and males is the same, with and without age adjustment. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for child gender, how many months pregnant the woman reported being at baseline, the mother's height and MUAC at baseline. Significance: * (10%); ** (5%); *** (1%).

Table 2.A24: New Child Health, by gender

Sample: In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	All		Males		Females		M-F Diff.
	C Mean	Effect	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of vaccinations	1.97 (1.48)	0.51*** (0.11)	1.93 (1.49)	0.56*** (0.13)	2.01 (1.48)	0.46*** (0.13)	0.605
Given deworming medication in past 6 months	17.12	9.65*** (2.19)	15.78	11.93*** (2.71)	18.69	7.68** (3.06)	0.298
Had illness/injury in past 30 days	72.47	−7.02*** (2.59)	73.06	−9.52*** (3.42)	71.78	−3.89 (3.30)	0.127
Had diarrhoea in past 2 weeks	38.06	−7.11*** (2.32)	40.67	−10.28*** (3.29)	34.97	−4.37 (3.25)	0.171

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. *Number of vaccinations* is an index that considers the number of vaccinations the child received among the following: BCG, polio, DPT, hepatitis B, yellow fever, and measles. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the subset of male children, and (5)-(6) for female children. Column (7) shows the p-value for the hypothesis that the estimated effects are equal across male and female children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.A25: Old Child Health, by gender

Sample: Older Sibling of In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	All		Males		Females		M-F Diff.
	C Mean (1)	Effect (2)	C Mean (3)	Effect (4)	C Mean (5)	Effect (6)	p (7)
Given deworming medication in past 6 months	20.30	10.47^{***} (2.92)	20.85	8.42^{**} (3.72)	19.82	11.97^{***} (3.67)	0.426
Had illness/injury in past 30 days	67.49	-8.02^{***} (3.10)	65.98	-6.78^{*} (3.85)	70.13	-9.14^{**} (3.90)	0.709
Had diarrhoea in past 2 weeks	20.88	-6.32^{**} (2.45)	21.76	-4.67 (3.12)	20.87	-8.58^{***} (3.16)	0.444

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the subset of male children, and (5)-(6) for female children. Column (7) shows the p-value for the hypothesis that the estimated effects are equal across male and female children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child. Significance: * (10%); ** (5%); *** (1%).

Table 2.A26: New Child antenatal care and breastfeeding practices, by gender

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

	All		Males		Females		M-F Diff.
	C Mean (1)	Effect (2)	C Mean (3)	Effect (4)	C Mean (5)	Effect (6)	p (7)
Practices index (using †)	−0.00 (1.00)	0.55*** (0.08)	0.00 (0.99)	0.55*** (0.09)	−0.00 (1.01)	0.54*** (0.10)	0.950
Prenatal							
Received antenatal care†	60.96	8.96** (3.86)	62.69	6.40 (4.32)	58.90	11.95*** (4.47)	0.175
Perinatal							
Fed colostrum in the first hour†	37.97	28.86*** (3.17)	36.62	30.02*** (3.76)	39.57	26.96*** (3.98)	0.352
Put to the breast immediately†	44.32	25.97*** (3.13)	43.42	26.14*** (3.71)	45.37	25.25*** (3.78)	0.646
Born at health facility†	12.78	5.48** (2.14)	12.44	7.51*** (2.58)	13.19	3.24 (2.61)	0.288
Postnatal							
Exclusively breastfed for 6-7 m†	11.53	30.13*** (2.97)	11.43	32.01*** (3.25)	11.66	28.50*** (3.69)	0.281

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. The index variable is computed using the methodology in Anderson (2008), and is standardised to have mean zero and variance one in the control group. The weights for the index are computed using the entire sample at midline. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the subset of male children, and (5)-(6) for female children. Column (7) shows the p-value for the hypothesis that the estimated effects are equal across male and female children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child and the child's age in months. Significance: * (10%); ** (5%); *** (1%).

Table 2.A27: New Child Dietary Diversity, by gender

Sample: In Utero New Child

ITT Estimates, Standard Errors Clustered by Village in Parentheses

	All		Males		Females		M-F Diff.
	C Mean	Effect	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dietary diversity index (using †)	3.28 (1.13)	0.35*** (0.07)	3.23 (1.16)	0.48*** (0.08)	3.33 (1.11)	0.21*** (0.08)	0.001
Food groups							
Grains, tubers, roots†	98.46	0.08 (0.65)	98.45	0.21 (0.88)	98.47	−0.05 (0.91)	0.493
Fruit and vegetables	85.53	1.37 (1.80)	86.01	1.72 (2.44)	84.97	1.21 (2.15)	0.787
Dark green leafy vegetables	43.26	−8.80*** (2.46)	43.52	−7.71** (3.26)	42.94	−9.25*** (3.32)	0.857
Vit-A rich fruit and veg.†	64.61	8.03*** (2.34)	65.80	8.00*** (3.00)	63.19	8.09** (3.15)	0.738
Other fruit and vegetables†	45.22	8.10*** (2.83)	44.56	11.17*** (3.52)	46.01	4.97 (3.54)	0.074
Nuts, beans, and seeds†	60.81	4.59* (2.48)	57.77	10.06*** (3.18)	64.42	−1.53 (3.27)	0.021
Animal-source foods	38.76	15.56*** (2.65)	36.79	18.10*** (3.33)	41.10	13.11*** (3.39)	0.187
Flesh foods (meat, fish)†	15.17	6.52*** (2.05)	14.51	7.90*** (2.43)	15.95	4.58* (2.56)	0.501
Eggs†	0.70	1.06** (0.44)	0.78	1.63** (0.63)	0.61	0.39 (0.53)	0.187
Milk, cheese, yogurt†	28.23	14.06*** (2.53)	27.20	15.75*** (3.43)	29.45	12.59*** (3.11)	0.275

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. The diet diversity index and the food groups are obtained from a 24-h food recall module administered to the child's mother or main carer. Each meal consumed in the day before the interview from waking up to bedtime is recorded, and each ingredient is coded into categories. The index uses slightly different food groups as the one presented below, according to WHO (2008). It is derived by summing the number of food groups the child has received, among the following categories: (i) Grains, roots and tubers; (ii) Legumes and nuts; (iii) Dairy products; (iv) Flesh foods; (v) Eggs; (vi) Vitamin-A rich fruits and veg.; (vii) Other fruits and vegetables. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the subset of male children, and (5)-(6) for female children. Column (7) shows the p-value for the hypothesis that the estimated effects are equal across male and female children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child and the child's age in months. Significance: * (10%); ** (5%); *** (1%).

Table 2.A28: Old Child Dietary Diversity, by gender
Sample: Older Sibling of In Utero New Child
ITT Estimates, Standard Errors Clustered by Village in Parentheses

	All		Males		Females		M-F Diff.
	C Mean	Effect	C Mean	Effect	C Mean	Effect	p
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dietary diversity index (using †)	3.53 (1.00)	0.26*** (0.06)	3.51 (0.97)	0.27*** (0.08)	3.55 (1.03)	0.25*** (0.09)	0.591
Food groups							
Grains, tubers, roots†	99.16	0.14 (0.50)	99.17	0.22 (0.72)	99.13	−0.03 (0.75)	0.838
Fruit and vegetables	94.13	−0.57 (1.26)	93.78	−0.94 (1.80)	94.81	−0.14 (1.92)	0.786
Dark green leafy vegetables	50.31	−8.51** (3.32)	52.28	−7.83** (3.93)	47.62	−8.00* (4.44)	0.750
Vit-A rich fruit and veg.†	72.54	6.33** (2.83)	70.54	10.02*** (3.69)	74.89	2.14 (4.00)	0.167
Other fruit and vegetables†	53.67	6.15* (3.17)	50.21	7.49* (4.02)	58.01	4.67 (4.06)	0.450
Nuts, beans, and seeds†	66.67	1.70 (3.07)	68.05	1.50 (3.81)	64.50	2.24 (4.23)	0.893
Animal-source foods	41.51	10.98*** (3.10)	41.08	10.87*** (4.02)	41.99	11.37*** (4.32)	0.692
Flesh foods (meat, fish)†	17.19	6.62*** (2.43)	17.01	8.73*** (2.99)	17.75	4.32 (3.12)	0.114
Eggs†	0.42	0.41 (0.34)	0.41	0.55 (0.52)	0.43	0.34 (0.44)	0.730
Milk, cheese, yogurt†	28.72	10.12*** (2.87)	28.63	7.54** (3.68)	28.57	12.84*** (4.04)	0.456

Notes: The main sample for this table is all households where the woman reported being pregnant at baseline and the new child is estimated to have been in utero at baseline ($N = 2,216$). The table reports ITT estimates of the effects of CDGP. Each row corresponds to a different outcome indicator. The diet diversity index and the food groups are obtained from a 24-h food recall module administered to the child's mother or main carer. Each meal consumed in the day before the interview from waking up to bedtime is recorded, and each ingredient is coded into categories. The index uses slightly different food groups as the one presented below, according to WHO (2008). It is derived by summing the number of food groups the child has received, among the following categories: (i) Grains, roots and tubers; (ii) Legumes and nuts; (iii) Dairy products; (iv) Flesh foods; (v) Eggs; (vi) Vitamin-A rich fruits and veg.; (vii) Other fruits and vegetables. Col. (1) shows the mean (and SD, if continuous) of the dependent variable in the control group at midline, for the New Child (born after baseline). Col. (2) shows the ITT effect. Standard errors are in parentheses (clustered at the village level). Columns (3)-(4) are repeated for the subset of male children, and (5)-(6) for female children. Column (7) shows the p-value for the hypothesis that the estimated effects are equal across male and female children. Means and effects on binary variables are reported in percentage points. All estimates control for region effects, randomisation tranche, and the following set of baseline covariates: total equivalised per capita household expenditure, age of the woman, whether she ever attended school, whether she is in a polygamous marriage, and number of household members in each of the following age groups: 0-2, 3-5, 6-12, 13-17, 18-64, 65+. In addition, we control for the gender of the child and the child's age in months. Significance: * (10%); ** (5%); *** (1%).

2.9.5 Appendix figures

Figure 2.A1: Instructional Materials Examples

Notes: Example of instructional materials from the programme curriculum. Source: CDGP facilitator guide.

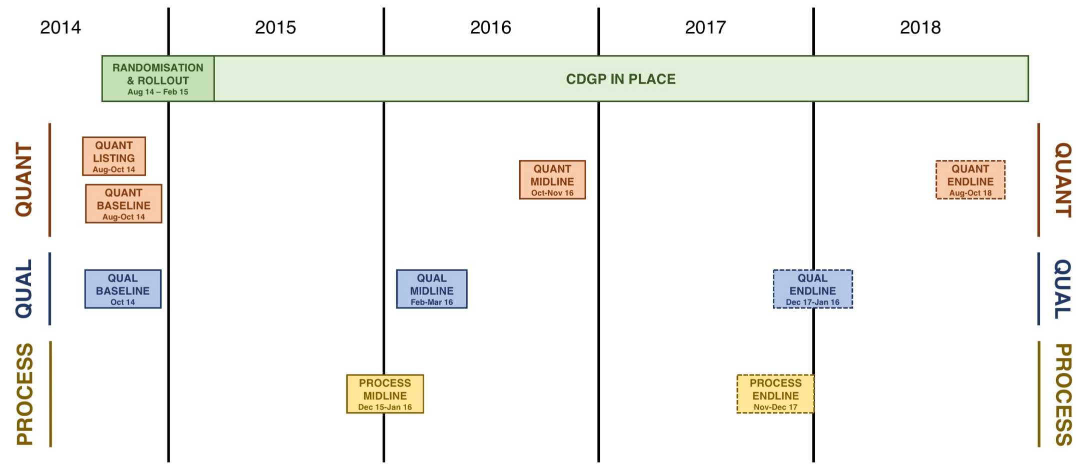

Figure 2.A2: CDGP Evaluation Timeline

Notes: The figure depicts a timeline of the evaluation process for CDGP, with the three work streams of the evaluation – quantitative, qualitative, and process.

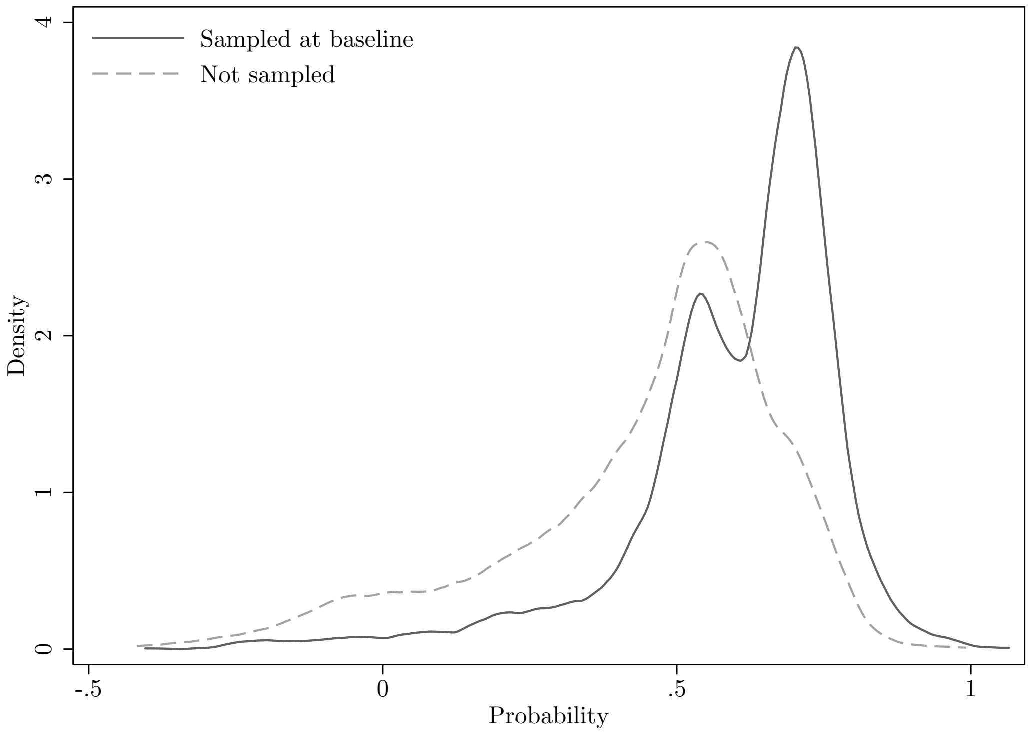

Figure 2.A3: Predicted pregnancy probability

Notes: The figure shows the distribution of predicted probabilities of becoming pregnant in the following two years, based on the listing exercise conducted before the main baseline data collection. The sample constitutes of all women aged 12-49 residing in evaluation villages who were not pregnant at the time of the listing interview. Pregnancy probabilities are plotted separately for women sampled at baseline and not sampled at baseline, under the assumption that the woman with the highest pregnancy probability was chosen in each household to be sampled at baseline. They are calculated using coefficients from a prediction model estimated on the Nigeria 2013 DHS dataset (NPC and ICF, 2014): the probability of giving birth in the next two years was modelled as a linear function of woman's age, time since last birth, household size, number of children aged under and over 5 years in household, and TV ownership. The distribution is plotted using kernel density estimation.

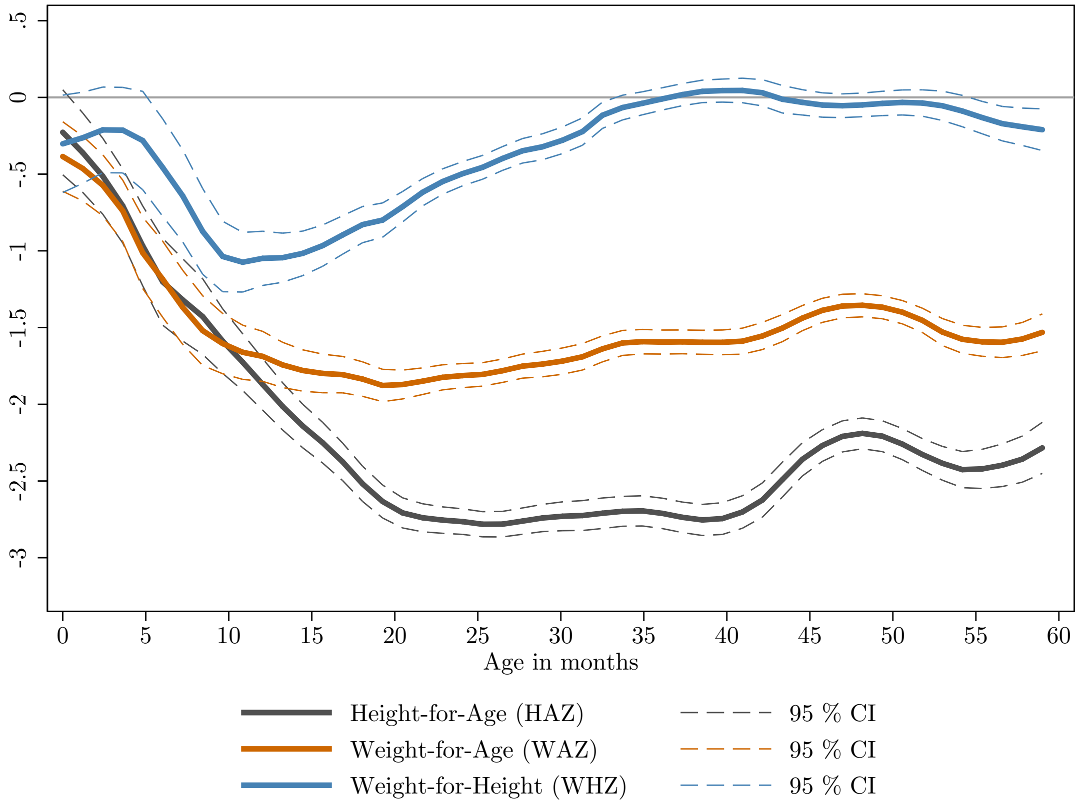

Figure 2.A4: Old Child anthropometrics - profiles of z-scores
Sample: Old children at baseline

Notes: The main sample for this graph is all households where a child aged 0-59 months was surveyed before the intervention, at baseline ($N = 4,164$). It shows smoothed age profiles of mean anthropometric z-scores obtained using a local mean kernel-weighted smoother.

Chapter 3

Inequality in socioemotional skills: a cross-cohort comparison

3.1 Introduction

Human capital is a key determinant of economic growth and performance and of the resources an individual creates and controls over the life cycle. Human capital is also important for various determinants of individual well-being, ranging from health to life satisfaction. In recent years, the process of human capital accumulation has received considerable attention. There is growing consensus on the fact that human capital is a multidimensional object, with different domains playing different roles in labour market as well as in the determination of other outcomes, including the process of human development. It is also recognised that human capital is the output of a very persistent process, where early years inputs play an important and persistent role.

And yet, there are still large gaps in our knowledge of the process of human capital development. These gaps are partly driven by the scarcity of high quality longitudinal data measuring the evolution over the life cycle of different dimension of human capital. Moreover, there is a lack of consensus on the best measures and on the tools to collect high quality data. As a consequence, even when data are available in different contexts, their comparability is problematic.

In this paper, we focus on an important dimension of human capital, which, so far, has received limited attention: socio-emotional skills. Evidence has shown that gaps in socio-emotional skills emerge at very young ages, and that in the absence of interventions are very persistent across the life cycle (Cunha et al., 2006). However, there is surprisingly little evidence on how inequality in this important dimension of human capital has changed across cohorts. In this paper, we start addressing this gap and focus on the measurement of these skills in two British cohorts: the one of children born in 1970 and the one of children born in 2000. We consider the measurement of socio-emotional skills during early childhood, as these skills have been shown, in a variety of contexts (Almlund et al., 2011) to have important long-run effects. Our goal is to characterise the distribution of socio-emotional skills in these cohorts and compare them. In the last part of the paper, we also consider the predictive power of different socio-emotional skills for health and socioeconomic outcomes.

The main contributions of the paper are four. First, we use two validated scales of childhood behavioural traits and select those items which are comparable across the two cohorts. By performing an exploratory factor analysis, we determine that we need at least two dimensions to characterise socio-emotional skills. In accordance with the child psychology tradition (Campbell, 1995), we label them as 'internalising' and 'externalising' skills, the former relating to the ability of children to focus their drive and determination and the latter relating to their ability in engaging in interpersonal activities.

Second, we study the comparability of the measures in the two cohorts. In particular, we test for *measurement invariance* of the items we use to estimate the latent factors. Intuitively, if one assumes that a set measures is related to a latent unobserved factor of interest, one can think of this relationship as being driven by the saliency of each measure and the level. If one uses a given measure as the relevant metric for the relevant factor, its saliency will determine the scale of the factor, while some other parameters, which could be driven by the difficulty of a given test or the social norms and attitudes towards a certain type of behaviour, determine the *average level* of the factor. Comparability of estimated factors across different groups (such as different cohorts) assumes that both the parameters that determine the saliency of a given set of measures and the level of the factors do not vary across groups. We find that, for the measures we use and for both factors, we cannot reject measurement invariance for the saliency parameters. However, we strongly reject measurement invariance for the level parameters. These results imply that while the level of inequality across the two cohorts in the skills we consider is comparable, we cannot determine whether the *average levels* of the two factors are larger or smaller in one of the two cohorts.

Third, given the results we obtain on measurement invariance, we proceed to compare the level of inequality in the two types of socio-emotional skills across the two cohorts, for both boys and girls. We find that the most recent cohort is more unequal in both dimensions of socio-emotional skills than the 1970 cohort. This result is particularly apparent for boys, and when looking at differences by maternal background. Fourth, we study whether the socio-emotional skills we observe at a young age are an important determinant of a variety of adolescent (and adult, for the older BCS cohort) outcomes. We find that socio-emotional skills at age five are more predictive than cognitive skills for unhealthy behaviours like smoking and measures of health capital such as body mass index. The effect of cognition, instead, dominates for educational and labour market outcomes.

The rest of the paper is organised as follows. In section 3.3, we briefly discuss the data we use in the analysis. In section 3.4, we discuss the methods we use to identify the number of dimensions in socio-emotional skills and how we estimate the latent factors that represent them. In section 3.5, we discuss the comparability of factors estimated with a given set of measures from different groups and the *measurement invariance* tests we use. Section 3.6 reports our empirical results, while section 3.7 concludes the paper.

3.2 Literature

The importance of cognition in predicting life course success is well established in the economics literature. However, in recent years the role played by 'non-cognitive' traits is being increasingly investigated. These traits include constructs as different as psychological and preference parameters such as social and emotional skills, locus of control and self-esteem, personality traits (e.g. conscientiousness), and risk aversion and time preferences. Given the vastness of this literature, we briefly review below the main papers on the determinants and consequences of socio-emotional traits which are more directly related to our work, and we refer to other sources for more exhaustive reviews (Borghans et al., 2008; Almlund et al., 2011; Goodman et al., 2015; Kautz et al., 2014).

Consequences of socio-emotional traits One of the first papers to show the importance of non-cognitive personality variables for wages was Bowles et al. (2001). Heckman et al. (2006) suggested that non-cognitive skills are at least as important as cognitive abilities in determining a variety of adults outcomes. Lindqvist and Vestman (2011), using data based on personal interviews conducted by a psychologist during the Swedish military enlistment exam, show that both cognitive and noncognitive abilities are important in the labour market, but for different outcomes: low noncognitive abilities are more correlated with unemployment or low earnings, while cognitive ability is a stronger predictor of wages for skilled workers. Segal (2013), using data on young men from the US National Education Longitudinal Survey, shows that eight-grade misbehaviour is important for earnings over and above eight-grade test scores. Layard et al. (2014) find that childhood emotional health (operationalised using the same mother-reported Rutter scale we use in the 1970 British cohort study) at ages 5, 10 and 16 is the most important predictor of adulthood life satisfaction and life course success.

There are only few studies in economics specifically studying "non-cognitive" traits and health behaviors. Conti et al. (2010) and Conti et al. (2011) were the first to consider three early endowments, including child socio-emotional traits and health in addition to cognition, using rich data from the 1970 British cohort study. They find strong evidence that non-cognitive traits promote health outcomes and healthy behaviors, and than not accounting for them overestimates the effects of cognition; additionally, they document that child cognitive traits are more important predictors of employment and wages than socio-emotional traits or early health. Chiteji (2010) used the US Panel Study of Income Dynamics (PSID) and found that future orientation and self-efficacy (related to emotional stability) are associated with less alcohol consumption and more exercise. Cobb-Clark et al. (2014) used the Australian HILDA data and found that an internal locus of control (also related to emotional stability, perceived control over one's life) is related to better health behaviours (diet, exercise, alcohol consumption anwhad smoking). Mendolia and Walker (2014) used the Longitudinal Study of Young People in England and found that individuals with external locus of control, low self-esteem, and low levels of work ethics, are more likely to engage in risky health behaviours. Savelyev and Tan (2019) show that the association between personality traits and health behaviours also holds in a high-IQ sample (the Terman Sample). Heckman et al. (2016) use, instead, early risky behaviours to measure socio-emotional traits, and confirm their predictive power for health behaviours and health outcomes.

Very few papers attempt to make cross-cohorts comparisons about the importance of socio-emotional skills. Blanden et al. (2007) – one of the closest study to ours – examine cognitive skills, non-cognitive traits, educational attainment and labour market attachment as mediators of the decline in inter-generational income mobility in UK between the 1958 and the 1970 cohorts. The authors take great care in selecting non-cognitive items to be as comparable as possible across cohorts, from the Rutter scale at age 10 for the 1970 cohort and from the Bristol Social Adjustment Guide for the 1958 cohort; however, they do not carry out a formal test of measurement invariance and the do not construct factor scores fully comparable across cohorts as we do. Another paper related to ours is the one by Reardon and Portilla (2016), who study recent trends in income, racial, and ethnic school gaps in several dimensions of school readiness, including academic achievement, self-control, and externalizing behavior, at kindergarten entry, using comparable data from the Early Childhood Longitudinal Studies (ECLS-K and ECLS-B) for cohorts born from the early 1990s to the 2000–2010 period in the US. They find that readiness gaps narrowed modestly from 1998 to 2010, particularly between high- and low-income students and between White and Hispanic students. Lastly, Deming (2017) uses a comparable set of skill measures and covariates across survey waves for the NLSY79 and the NLSY97, and finds that the labour market return to social skills was much greater in the 2000s than in the mid-1980s and 1990s.

Determinants of socio-emotional traits Equally flourishing has been the literature on the determinants of child socio-emotional skills, which ranges from reduced-form, correlational or causal estimates, to more structural approaches. One of the first papers by Segal (2008) has shown that a variety of family and school characteristics predict classroom behaviour. Carneiro et al. (2013) study the intergenerational impacts of maternal education, using data from the NLSY79 and an instrumental variable strategy; they find strong effects in terms of reduction in children's behavioural problems. Cunha et al. (2010) and Attanasio et al. (2018) both estimate production functions for child cognitive and socio-emotional development (in US and Colombia, respectively), and find an important role played by parental investments.

Interventions for improving Social and Emotional Learning (SEL) in a school setting have shown significant improvements in socio-emotional skills, attitudes, behaviours, and academic performance (Durlak et al., 2011), and a substantial positive return on investments (Belfield et al., 2015); after-school programs have been shown to be equally effective (Durlak et al., 2010).

Additionally, it has been shown that a key mechanism through which early childhood interventions improve adult socioeconomic and health outcomes is by boosting socio-emotional skills, such as four teacher-reported behavioural outcomes in the project STAR[1] (Chetty et al., 2011), reductions in externalising behaviour (from the Pupil Behavior Inventory) at ages 7-9

[1] Student's effort, initiative, non-participatory behavior, and how the student is seen to 'value' the class.

in the Perry Preschool Project (Heckman et al., 2013; Conti et al., 2016), or improvements in task orientation at ages 1-2 in the Abecedarian Project (Conti et al., 2016).

In sum, even if the literature on the determinants and consequences of socio-emotional skills has been booming, most papers use skills measured in late childhood or in adolescence; and no paper in economics formally tests for invariance of measurements across different groups and constructs fully comparable scores. In this paper, we use measures of child socio-emotional development at age 5, hence at the very beginning of formal schooling; and we construct comparable scales across the two cohorts we study (the 1970 and the 2000 British cohorts), so that we can investigate changes in inequality in early development, their determinants, and consequences, in a parallel fashion.

3.3 Data

We use information from two nationally representative longitudinal studies in the UK. The studies follow the lives of children born approximately 30 years apart: the British Cohort Study (BCS) surveys individuals born in 1970, and the Millennium Cohort Study (MCS) includes births between 2000 and 2002. The British Cohort Study includes all individuals born across Great Britain in a single week in 1970. Cohort members' families – and subsequently members themselves – were surveyed on multiple occasions. For this paper we augment the information at the five-year survey with data from birth, adolescence (16), and adulthood (30, 38, 42). The Millennium Cohort Study follows individuals born in the UK between September 2000 and January 2002. We use the first (age 0) survey – carried out at 9 months of age – and the sweeps at around 5 and 14 years of age.[2]

Our main focus is on socio-emotional skills of children around age five. We take advantage of the longitudinal nature of the cohorts by merging information from surveys before and after age five. From the birth survey, we include information on gestational age and weight at birth, previous stillbirths, parity, maternal smoking in pregnancy, maternal age, height, and marital status. From the five year survey, we extract maternal education, employment status, and the father's occupation. All the above variables are transformed or recoded to maximise comparability between the two studies. Furthermore, we add some adolescent outcomes such as smoking and BMI, with the caveat that these are surveyed at different ages – 16 in BCS and 14 in MCS. Finally, for the 1970 cohort we also include measures of adult educational attainment, BMI, and income. Variable definitions are available in Table 3.A2.

Ideally, we would compare socio-emotional skills alongside cognitive skills. However, the cognitive tests administered to each cohort have no overlap, even at the item level. We thus use the available cognitive tests in each cohort to estimate simple confirmatory factor model with a single latent dimension, separately by cohort (see Table 3.A2 for the tests used). Unlike the other indicators in our analysis, cognitive skills are thus not comparable across cohort.

[2] All data is publicly available at the UK Data Service (Chamberlain, 2013; Butler, 2016a,b, 2017; University Of London. Institute Of Education. Centre For Longitudinal Studies, 2016a,b,c, 2017a,b,c).

Another complication arises from the fact that, differently from the British Cohort Study, the Millennium Cohort Study has a stratified design. It oversamples children living in administrative areas characterised by higher socioeconomic deprivation and larger ethnic minority population (Plewis et al., 2007). We rebalance the MCS sample to make it nationally representative by excluding from the analysis a fraction of observations from the oversampled areas, proportionally to their sampling probability.[3] We also restrict our sample to individuals born in England. Finally, we restrict the sample to cases where the respondent in the five-year followup was the natural mother, and where there is complete information on socioemotional skills. The final sample contains 9,545 individuals from the British Cohort Study, and 5,436 from the Millennium Cohort Study.

3.4 Dimensions of socio-emotional skills

Child socio-emotional skills are an unobservable and difficult to measure construct. Over recent years, the measurement of such skills has evolved and, over time, different measures have been used. As we discuss below, this makes the comparison of socio emotional skills across different groups, assessed with different tools, difficult.

A common approach to infer a child's socio-emotional development is based on behavioural screening scales. As part of these tools, mothers (or teachers) indicate whether their children exhibit a series of behaviours – the *items* of the scale. In the British and Millennium Cohort Studies, two different scales were employed. In the BCS, the Rutter A Scale was used (Rutter et al., 1970) while in the MCS cohort, mothers were administered the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1994, 1997). The SDQ was created as an update to the Rutter scale. It encompasses more recent advances in child psychopathology, and emphasises positive traits alongside undesirable ones (Stone et al., 2010). The Rutter and SDQ scales are reproduced in Table 3.A1; they have 23 and 25 items each, respectively. In the child psychiatry and psychology literatures, the Rutter and SDQ behavioural screening scales are regarded as measures of behavioural problems and mental health. However, in our analysis we follow the economics literature, and - after having recoded them accordingly - we interpret them as measures of positive child development (Goodman and Goodman, 2011).

While the Rutter and SDQ scales are similar in their components, there is no a priori reason to expect them to be directly comparable. First, the overlap of behaviours described in the two scales is only partial. Second, the wording of each item is slightly different, both in the description and in the options that can be selected as answers. Third, the different ordering of the items within each scale might lead to order effects. Fourth, and no less importantly, the interpretation of each behaviour by respondents living 30 years apart (1975 vs 2006) might differ due to a host of evolving societal norms.

As our goal is to compare socio-emotional skills across the two cohorts, we construct a

[3] See Table 5.5 in Plewis et al. (2007). This choice is mainly driven by software limitations. The `lavaan` package in **R** (Rosseel, 2012) is the most suitable tool for our invariance analysis, but it does not allow to use weights when outcomes are categorical, as it is the case for the socio-emotional measurements.

new scale by retaining the items that are worded in a similar way across the two original Rutter and SDQ scales, and making some slight coding adjustments to maximise comparability. In what follows, we will consider the included items to be the same *measure* in the two cohorts. The wording of the items we will be using in the analysis is presented in Table 3.1: we retain 13 items for the BCS (two of them are grouped) and 11 for the MCS with high degree of comparability.[4]. Item-level prevalence by cohort and gender is in Table 3.A4. We see that, in general, item prevalence is more similar across genders within the same cohort, than across cohorts. For the majority of items, there is a lower prevalence of problematic behaviours in the MCS than in the BCS; however, four items (distracted, tantrums, fearful, aches) show a higher prevalence in 2006 than in 1975. Regardless, a simple cross-cohort comparison of item-level prevalence is misleading because of changing perceptions and norms about what constitutes problematic behaviour in children. The analysis in section 3.5 tackles this issue.

In the remainder of this section, we analyse the properties of the new scale. In particular, we study the *factor structure* of our scale. Namely, we establish how many latent dimensions of socio-emotional skills the scale is capturing, and which items are measuring which dimension. We then estimate the parameters of the factor models that corresponds to our choice of dimension and attribution of specific items to factors. In the following section, we investigate to what extent socio-emotional skills are measured in the same way across cohorts.

3.4.1 Exploratory analysis

The original Rutter scale, used for the BCS cohort, distinguishes behaviours into two subscales: *anti-social* and *neurotic* (Rutter et al., 1970). This two-factor conceptualisation has been validated using data from multiple contexts, and the latent dimensions have been broadly identified as externalising and internalising behaviour problems.[5] The Strength and Difficulties Questionnaire, used for the MCS cohort, was conceived to have five subscales of five items each. The five subscales are: *hyperactivity*, *emotional symptoms*, *conduct problems*, *peer problems*, and *prosocial*. This five-factor structure has been validated in many contexts (Stone et al., 2010); lower-dimensional structures have been also suggested (Dickey and Blumberg, 2004). Recent research has shown that there are some benefits to using broader subscales that correspond to the externalising and internalising factors in Rutter, especially in low-risk or general population samples (Goodman et al., 2010).

We use exploratory factor analysis (EFA) to assess the factors structure of our 11-item scale combining Rutter and SDQ.[6] We start by investigating the number of latent constructs

[4]We exclude from the analysis items that were completely different between the two questionnaires, although we could have included them in the factor analysis and treated them as missing in the cohort were they were not administered. While this could have improved efficiency, we decided to rely on a more coherent set of measures to maximise comparability between the two cohorts.

[5]See for example Fowler and Park (1979); Venables et al. (1983); Tremblay et al. (1987); Berglund (1999); Klein et al. (2009). However, in some cases a three-factor structure was found to better fit the data, with the externalising factor separating into two factors seemingly capturing aggressive and hyperactive behaviours (Behar and Stringfield, 1974; McGee et al., 1985).

[6]Factor-analytic methods have long been used in psychology, and in recent years they have become increasingly popular in economics, especially to meaningfully aggregate high-dimensional items measuring different

that are captured by the scale, using different methods developed in the psychometric literature, and now also used in the economics literature. The results are displayed in Table 3.A6. As pointed out in Conti et al. (2014), there is relatively little agreement among procedures; this is the case especially for the Rutter items in the BCS data, where different methods suggest to retain between 1 and 3 factors, while most methods suggest to retain 2 factors in the MCS. In our analysis, we adopt two factors and a dedicated measurement system, where each measure reflects only one factor. This choice is justified both by the child psychology literature cited above, and as compromise to work with the same number of factors in the two cohorts.

The two-factor EFA delivers a neat and sensible separation between items, as shown in Table 3.A7: reassuringly, similarly-worded items load on the same factor across the two cohorts, and also the magnitude of the respective loadings (measuring the strength of the association between the item and the factor) is very smilar. Following naming conventions from previous research in child psychology (Campbell, 1995), we name the first dimension *Externalising skills* (EXT, indicating low scores on the items restless, squirmy/fidgety, fights/bullies, distracted, tantrums, and disobedient) and the second dimension *Internalising skills* (INT, indicating low scores on the items worried, fearful, solitary, unhappy, and aches).

3.4.2 Factor model

Equipped with the factor structure inferred in the previous section, we specify a multiple-group factor analysis model to formally quantify the strength of the relationship between the observed items in our scale and the two latent socio-emotional skills. We specify two groups of children $c = \{BCS, MCS\}$, corresponding to the two cohorts. Each individual child is denoted by $j = 1 \dots N_c$, where N_c is the number of children in cohort c. For each child j in cohort c, we observe categorical items X_{ijc} with $i = 1, \dots, 11$, corresponding to the eleven maternal reports in Table 3.1. We assume that each child is characterised by a latent bi-dimensional vector of externalising and internalising socio-emotional skills $\boldsymbol{\theta}_{jc} = (\theta_{jc}^{EXT}, \theta_{jc}^{INT})$, as shown by the EFA in the previous section.

Children are assumed to have a latent continuous propensity X_{ijc}^* for each item $i = 1, \dots, I$. We model this propensity as a function of item- and cohort-specific intercepts ν_{ic} and loadings $\boldsymbol{\lambda}_{ic}$, and the child's latent skills $\boldsymbol{\theta}_{jc}$, plus an independent error component u_{ijc}. The propensity for each item can be written as follows:

$$X_{ijc}^* = \nu_{ic} + \boldsymbol{\lambda}_{ic}\boldsymbol{\theta}_{jc} + u_{ijc} \qquad \text{for} \quad i = 1, \dots, 11$$

or more compactly:

$$\boldsymbol{X}_{jc}^* = \boldsymbol{\nu}_c + \boldsymbol{\Lambda}_c\boldsymbol{\theta}_{jc} + \boldsymbol{u}_{jc} \tag{3.4.1}$$

We make the common assumption of a dedicated (or congeneric) factor structure, where each measure is assumed to load on only one latent dimension (Heckman et al., 2013; Conti

aspects of common underlying dimensions of human development. The EFA is performed decomposing the polychoric correlation matrix of the items and using weighted least squares, and the solution is rescaled using oblique factor rotation (*oblimin*). We use the **R** package `psych`, version 1.8.4 (Revelle, 2018).

et al., 2010; Attanasio et al., 2018). We mirror the structure found in the exploratory factor analysis (see Table 3.A7), and assume that items 1-6 load exclusively on the externalising factor and items 7-11 on the internalising factor.[7]

The discrete ordered nature of the observed measures X_{ijc} is incorporated by introducing item- and cohort-specific threshold parameters τ_{ic} (Muthén, 1984). The observed measures as a function of the propensities X^* can be then written as follows:

$$X_{ijc} = s \qquad \text{if } \tau_{s,ic} \leq X^*_{ijc} < \tau_{s+1,ic} \qquad \text{for } s = 0, 1, 2 \tag{3.4.2}$$

with $\tau_{0,ic} = -\infty$ and $\tau_{3,ic} = +\infty$. Notice that we recode all ordered items to have higher values for *better* behaviours, so that our latent vectors can be interpreted as favourable skills and not behavioural problems.

The model implies the following expression for the mean and covariance structure of the latent propensities:

$$\boldsymbol{\mu}_c = \boldsymbol{\nu}_c + \boldsymbol{\Lambda}_c \boldsymbol{\kappa}_c \qquad \text{and} \qquad \Sigma_c = \boldsymbol{\Lambda}_c \boldsymbol{\Phi}_c \boldsymbol{\Lambda}'_c + \boldsymbol{\Psi}_c.$$

The model restrictions in (3.4.1) and (3.4.2) do not identify the parameters without additional assumptions. As per the traditional factor analysis approach, we impose a normal distribution on the latent skills and error terms:[8]

$$\boldsymbol{\theta}_{jc} \sim N(\boldsymbol{\kappa}_c, \boldsymbol{\Phi}_c) \qquad \text{and} \qquad \boldsymbol{u}_{jc} \sim N(\mathbf{0}, \boldsymbol{\Psi}_c). \tag{3.4.3}$$

Even with these assumptions, there are infinite equivalent parameterisations through which the model can be identified – the well-known issue of factor indeterminacy. We follow common practice and identify the model by setting the mean $\boldsymbol{\kappa}$ and variance $\boldsymbol{\Phi}$ of the latent skill factor in both cohorts to zero and one, respectively. Furthermore, we set intercepts to zero and error variances to one. Loadings λ and thresholds τ are instead allowed to vary across cohorts.

$$\text{diag}(\boldsymbol{\Phi}_c) = \boldsymbol{I}, \quad \boldsymbol{\kappa}_c = \mathbf{0}, \quad \boldsymbol{\nu}_c = \mathbf{0}, \quad \text{and } \text{diag}(\boldsymbol{\Psi}_c) = \boldsymbol{I} \qquad \forall c \in \{BCS, MCS\}. \tag{3.4.4}$$

The restrictions in (3.4.1), (3.4.2), (3.4.3), and (3.4.4) define the so-called *configural* model. This is a 'minimum' identifiable model, in that it places the least possible restrictions on how parameters are allowed to vary across cohorts. It serves as a basis for our measurement invariance analysis in the next section.[9]

[7] The dedicated factor structure corresponds to a sparse loading matrix, i.e.:

$$\boldsymbol{\Lambda}_c := \begin{bmatrix} \lambda_{1c}, \ldots, \lambda_{6c} & \mathbf{0} \\ \mathbf{0} & \lambda_{7c}, \ldots, \lambda_{11c} \end{bmatrix}.$$

[8] Recent work has also used mixtures of normals for the latent factors distribution, e.g. Conti et al. (2010).

[9] This set of identifying restrictions is known as *Theta* parameterisation (Wu and Estabrook, 2016). See Appendix 3.10.2 for statistically equivalent alternative parameterisations.

3.5 Measurement invariance

Any comparison between socioemotional skills across the two cohorts requires that the measures at our disposal have the same relationship with the latent constructs of interest in both cohorts. In other words, the items in our new scale must measure externalising and internalising socioemotional skills in the same way in the BCS and MCS data. This property is denominated measurement invariance (Vandenberg and Lance, 2000; Putnick and Bornstein, 2016).

In the framework of factor analysis, measurement invariance is a formally testable property. In this paper, we follow the recent identification methodology by Wu and Estabrook (2016). The configural model defined in the previous section serves as the starting point. Measurement invariance is then assessed by comparing the configural model to a series of hierarchically nested models. These models place increasing restrictions on the item parameters, constraining them to be equal across groups. Their fit is then compared to that of the configural model. Intuitively, if the additional cross-group restrictions have not significantly worsened model fit, one can conclude that a certain level of invariance is achieved. The hierarchy of restrictions is detailed in Table 3.A3.

Let's consider examples from our application. A *loading and threshold invariance* model restricts every item's loading λ and threshold τ parameters to have the same value in the two cohorts. It assumes that the items in our scale have the same relationship with latent skills across the two cohorts. In other words, items have the same salience, or informational content relative to skills. If this model fits as well as the configural model, we can be confident that the socioemotional skills of children in the two cohorts can be placed on the same scale, and their *variances* can be compared. To see why, consider equation (3.4.1). If the loading matrix $\boldsymbol{\Lambda}$ is the same across cohorts, any difference in latent skills $\Delta\boldsymbol{\theta}$ will correspond to the same difference in latent propensities $\Delta\boldsymbol{X}^*$. Equality of thresholds τ ensures that propensities $\boldsymbol{X}^*$ map into observed items $\boldsymbol{X}$ in the same way.

A *loading, threshold, and intercept invariance* model additionally restricts every item's intercept ν across cohorts. A good relative fit of this model indicates that socioemotional skills can be compared across cohorts in terms of their *means* as well. To see why, consider the following. Since the λ and ν parameters are the same across cohorts, a child in the BCS cohort with a given level of latent skills $\bar{\boldsymbol{\theta}}$ will have the same expected latent item propensities $\boldsymbol{X}^*$ as a child with the same skills in the MCS cohort. Again, equality of thresholds τ fixes the mapping between $\boldsymbol{X}^*$ and $\boldsymbol{X}$.[10]

We estimate the sequence of models detailed in Table 3.A3 by Weighted Least Squares.[11]

[10] We recognise that simultaneous invariance of *all* items is not the minimum requirement for comparability. In theory, the availability of just one invariant item (known as 'anchor') would suffice to fix the scale and location of the system. However, partial invariance approaches are hard to implement in practice. Its validity hinges on selecting one (or more) truly invariant anchor, which is challenging on an a priori basis. The full procedure, restricting all parameters of a certain type across groups, does not identify which items are at the source of the invariance. Algorithms have been proposed to deal with this issue (Yoon and Millsap, 2007; Cheung and Lau, 2012), however there are still doubts on their robustness and their applicability to the categorical case (Vandenberg and Morelli, 2016).

[11] Parameters are estimated by mean- and variance-adjusted weighted least squares (WLSMV) – see Muthen

For the purposes of the analysis, we define groups c as cohort-gender cells, with the reference group being males in the BCS cohort. We then compare the fit of each model against the configural model.

Comparison of χ^2 values across models is a common likelihood-based strategy. However, tests based on $\Delta\chi^2$ are known to display high Type I error rates with large sample size and more complex models such as our own (Sass et al., 2014). In fact, for all invariance levels in our applications a chi-squared difference would point to a lack of measurement invariance. The use of approximate fit indices (AFIs) is therefore recommended alongside χ^2. These indices do not have a known sampling distribution, thus making it necessary to rely on rules of thumb to assess what level of $\triangle$AFI indicates invariance. Nevertheless, AFIs are widely used in empirical practice to assess model fit.[12]

The fit of each model is compared in Panel A of Table 3.A8. The model with restricted thresholds and loadings exhibits a comparable fit to the configural model, according to all the AFIs. Invariance of loadings and thresholds across cohorts implies that items in our scale are equally salient in their informational content, and that the latent propensities have equal mapping into the observed items. However, further restricting intercepts results in a model where invariance is rejected across the board.[13] In other words, intercept parameters in our model (ν) are estimated to be different between maternal reports in the British and Millennium Cohort Studies. This means that, for a given level of latent skills, mothers in MCS tend to assess behaviours differently from mothers in BCS. Thus, cohort differences in scores on our scale cannot be unequivocally interpreted as differences in the underlying skills, since they might also reflect differences in reporting.

This is an important finding, which has to our knowledge never been acknowledged in this literature. How can this lack of comparability be explained? A possible interpretation is

et al. (1997); estimation starts from the items' polychoric correlation matrix, uses diagonally weighted least squares (DWLS), and exploits the full weight matrix to compute robust standard errors and test statistics. Robust WLS has proved in simulation studies to be moderately robust to small violations of the normality assumption in the latent underlying measures (Flora and Curran, 2004), and generally outperforms maximum likelihood in large samples (Beauducel and Herzberg, 2006; Li, 2016). All estimates are computed using the `lavaan` package (version 0.6-2) in **R** (Rosseel, 2012).

[12]The root mean squared error of approximation (RMSE) and the Tucker-Lewis index (TLI) are traditionally the most used AFIs in empirical practice. Simulation evidence by Cheung and Rensvold (2002) shows that these indices can show correlation between overall and relative fit, and suggest relying on additional indices, such as the comparative fit index (CFI, Bentler, 1990), McDonald non-centrality index (MFI, McDonald, 1989), and Gamma-hat index (Steiger, 1989) for the case of ordered measures. Commonly accepted thresholds for rejection are $\Delta CFI < -0.01$, $\Delta MFI < -0.02$, and $\Delta Gammahat < -0.001$. Meade et al. (2008), using the results from a simulation study, suggests stricter thresholds that should apply in a variety of conditions. For CFI, a single cutoff value of .002 is proposed, while cutoffs for MFI depend on the problem's characteristics; in our case (2 factors, 11 items), they suggest .0066. Sass et al. (2014) however cast some doubts of the generalisability of these cutoffs to WLSMV estimators.

[13]We do not present fit results for the threshold-only invariance model, as it is statistically equivalent to the configural model and thus its fit is mathematically the same – see Table 3 in Wu and Estabrook, 2016. The ages at which socio-emotional skills are observed varies slightly between BCS and MCS, due to different sampling and fieldwork schedules. In the MCS cohort, the age distribution has significantly higher variance. In Panel B of Table 3.A8, we restrict the sample to 59 to 61 months, where the overlap between BCS and MCS is maximised. In Panels C and D, we restrict to male and female children respectively. In all these cases, invariance of thresholds and loadings is achieved, but invariance of intercepts is rejected. We can thus rule out that the lack of intercept invariance comes from differences in ages or invariance across child gender.

connected with secular evolution of social and cultural norms about child behaviours. For example, commonly held views of what constitutes a restless, distracted, or unhappy child might have changed between 1975 and 2006.[14]

To summarise, our measurement invariance analysis shows partial comparability of socioemotional skills across cohorts. In particular, the variance of skills can be compared across cohorts, but mean cohort differences do not necessarily reflect differences in skills. We can use scores from our scale to compare children within the same cohort, but not across cohorts. However, we can also compare within-cohort differences between groups of children, across cohorts. As an example, consider two groups of children A and B in the BCS cohort, and two groups of children C and D in the MCS. We cannot compare the mean level of skills between groups A and C, but we can compare the mean difference between groups A and B with the mean difference between groups C and D. This is the approach we take for the rest of the paper. Refraining from direct cross-cohort comparisons, we interpreting significance and magnitude of within-cohort differences across the cohorts.

3.6 Results

Parameter estimates from our factor model are presented in Table 3.A9. As discussed in the previous section, loadings and thresholds are constrained to have the same value across groups. Intercepts are normalised to zero, and error variances to one, for the reference group – males in the BCS cohort. We use the estimates from this model to predict a score for each child in our sample along the latent externalising and internalising socio-emotional skill dimensions.[15] We plot the distribution of the scores in Figure 3.1. The unit of measurement is standard deviations of the distribution in the subsample of males in the BCS. Given our measurement invariance results in section 3.5, we stress that the *location* of these scores should not be directly compared across cohorts. However, the shape of the distribution can be given a cross-cohort interpretation.[16] It is immediately visible that there is more mass in the tails of the distribution in the 2000 than in the 1970 cohort.

[14]Calibrating the Rutter and SDQ using a contemporary sample of children cannot rule out this issue. For example, Collishaw et al. (2004) administered both Rutter and SDQ items to parents of a small sample of adolescents in London. They use the mapping between the two questionnaires to impute Rutter scores for mothers who answered the SDQ. This can correct for contemporaneous reporting differences between questionnaires, but cannot tackle reporting differences between samples collected at different times in history.

[15]We use an empirical Bayes modal (EBM) approach to estimate the scores. The parameters are estimated using three sources of information. The first is the distribution of the latent variables θ, treated as random parameters with a prior $h(\theta, \Omega)$, conditional on the parameters Ω. This prior is assumed to be multivariate normal. The second is the observed data X, and the third is the estimated parameters $\hat{\Omega}$. Data and prior are combined into the posterior distribution $w(\theta|X, \hat{\Omega})$. For further details, see Chapter 7 in Skrondal and Rabe-Hesketh (2004).

[16]The density of the scored factors can be contrasted with the distribution of sum scores in Figure 3.1. Using raw scores, instead, shows an increase in mass only at the top of the distribution.

3.6.1 Inequality in socioemotional skills

We find that, both unconditionally and for specific groups, inequality in socio-emotional skills at age five has increased between 1975 and 2005/6. Table 3.2 shows unconditional inequality statistics, using quantile differences in the distribution of skills by gender and cohort. With the exception of internalising skills in female children, all distributions have widened substantially between the BCS and MCS cohorts. The gap for both externalising and internalising skills between the 90th and the 10th percentiles for males has increased by approximately half a standard deviation. The increase in the gap is more pronounced in the bottom half of the distribution. For females, we see a narrowing at the top (90-50), but a widening at the bottom (50-10) of the distribution, again for both externalising and internalising skills.

Inequality has also increased conditional on socioeconomic status. Figure 3.2 shows mean skills by maternal education. We compare mothers who continued education past the compulsory age with mothers who left school at the compulsory leaving age, according to their year of birth. Given lack of comparability in the level of skills across cohort, we normalise the mean in the 'Compulsory' group to zero for both cohorts. For both males and females, and for both externalising and internalising skills, the difference in the socio-emotional skills of their children between more and less educated mothers has increased. The size of the increase is around .1 to .15 of a standard deviation. The increase is particularly pronounced for males, for whom it goes from .20 to .30 for externalising and from .12 to .24 for internalising.

Figure 3.3 shows an even starker pattern when comparing children of mothers who smoked in pregnancy with non-smoking mothers. The fact that maternal smoking during pregnancy is a risk factor for offspring behavioural problems is well known in the medical literature (Gaysina et al., 2013); there is less evidence, however, on whether and to which extent these associations have changed across cohorts. The difference in child skills has increased, from less than .2 to around .4 of a standard deviation, again with the biggest increase experienced by the boys. There is also a significant increase in the gradient by paternal occupation based on social class (Figure 3.4), although this is less pronounced if compared to the one based on maternal characteristics. In particular, male children with no father figure living in their household have worse skills compared to children with blue collar fathers in the MCS cohort. Otherwise, skill differences in father's occupation are mostly constant across the two cohorts.[17] These patterns are in stark contrast with the findings of Reardon and Portilla (2016) for the US, who have found a narrowing of the readiness gaps from 1998 to 2010.

We then examine the same patterns as in the previous figures, but conditional on other family background indicators. The aim is to disentangle the relative contribution of each indicator to socio-emotional skills, and how it has changed in the thirty years between the two cohorts. Table 3.3 shows coefficients from linear regressions of socio-emotional skills at five on contemporaneous and past socioeconomic indicators, by cohort and gender. Coefficients

[17] Figures 3.2, 3.3, and 3.4 show inequality in the scale items underlying the factor scores used in this section. The increase in inequality across cohorts is still present, but less marked when looking at these single items. This shows the importance of the factor analysis step in aggregating items, explicitly modelling the measurement error, and testing and accounting for (loadings and thresholds) invariance across the two cohorts.

for indicators in BCS and MCS are presented side by side, together with the *p*-value of the hypothesis that coefficients are the same in the two cohorts.[18]

Overall, the importance of maternal socioeconomic status (education and in particular employment) in determining socio-emotional skills has increased from the BCS to the MCS children. The 'premium' in skills for children of better educated and employed mothers is significantly larger, for both boys and girls, internalising and externalising skills. At the same time, the penalty for having a blue-collar father, or not having a father figure at all in the household, has significantly declined across the two cohorts, especially for girls. Being born to an unmarried mother, and to a mother who smoked during pregnancy, is associated with a higher penalty for both dimensions of socio-emotional skills in the latter cohort, but only for males. This is consistent with recent evidence which shows that family disadvantage disproportionately impedes the pre-market development of boys, in terms of higher disciplinary problems, lower achievement scores, and fewer high-school completions (Autor et al., 2016).[19] Girls of non-white ethnicity, instead, have worse internalising and externalising skills in the MCS, a penalty not suffered by 5-year old non-white girls in the BCS. Firstborn boys and girls in the BCS have worse skills, but this difference disappears in the MCS. Lastly, we document an increase in the returns to birth weight, which is more pronounced for boys' internalising skills.

These changes in the relative importance of pregnancy factors and family background characteristics for child socio-emotional skills at age 5 need to be interpreted in the light of the significant changes in the prevalence of such characteristics across cohorts. As shown in Table 3.A5, the age of the mother at birth, proportion of mothers non-smoking in pregnancy, with post-compulsory education and in employment at the age 5 of the child has substantially increased; at the same time, the proportion of households with no father figure has increased, and so the proportion of women unmarried at birth is much higher in the 2000 than in the 1970 cohort. Also, the ethnic structure of the population has changed, with a higher proportion of non-white children in the MCS than in the BCS. In general, this has been a period of significant societal changes, with an almost continual rise in the proportion of women in employment, an older age at first birth and a rise in dual-earning parents families (Roantree and Vira, 2018). However, here we do not attempt to disentangle whether and to which extent the observed changes in inequality in socio-emotional skills across the two cohorts can be attributed to changes in returns (or penalties) to maternal characteristics (such as education and employment) or to compositional changes, as has been done for the analysis of wage inequality (Blundell et al., 2007).

[18] We also estimated Tobit models to account for the right truncation of the distribution of skills – see Figure 3.1. Tobit estimates are extremely similar to the linear estimates in Table 3.3, and are available from the authors upon request.

[19] It is important to underscore that there has been a significant rise is cohabitation between 1975 and 2006. It is likely that unmarried mothers in the two cohorts have very different characteristics. The choice of this indicator is due to the absence of information on cohabitation in the birth survey for the BCS cohort.

3.6.2 Socio-emotional skills and adolescent/adult outcomes

In this last section, we study the predictive power of socio-emotional skills for adolescent and adult outcomes. We contribute to a vast interdisciplinary literature by examining medium- and long-term impacts of skills measured at an earlier age than usually examined in previous studies, well before the start of formal education. We do so by regressing health and socioeconomic outcomes measured in adolescence and adulthood on the socio-emotional skills scores at age five obtained by our factor model, controlling for the harmonised family background variables at birth and age five (see Table 3.A2). We present results with and without controlling for cognitive skills. As detailed in Section 3.3, the available cognitive measures are not comparable across cohorts. Still, we control for a factor score that summarises all information on cognitive skills that is available in each cohort, regardless of their comparability.

Socio-emotional skills at five years of age are predictive of adolescent health behaviour and outcomes in both cohorts.[20] Table 3.4 examines adolescent smoking and BMI for both cohorts; Table 3.A10 reports the results for the same outcomes in adulthood (at age 42), for the BCS only. Externalising skills are negatively correlated to subsequent smoking and BMI in both cohorts, for both genders. Recall that a child with high externalising skills exhibits less restless and hyperactive behaviours, and has less anti-social conduct. Our findings are consistent with the body of evidence reviewed in section 3.2, which shows that better socio-emotional skills (measured using different scales and at various points during childhood and adolescence) are negatively associated with smoking. At the same time, internalising skills are positively correlated with smoking (only in the 1970 cohort) and BMI (only for girls), although less strongly than externalising skills. This apparently counterintuitive result makes sense in light of the items in our internalising scale shown in Table 3.1. A child with better internalising skills is less solitary, neurotic, and worried. From this perspective, she is likely more sociable and subject to peer influence in health behaviours. This is consistent with the evidence in Goodman et al. (2015), who find a positive association between child emotional health (measured with items for the internalizing behaviour subscale of the Rutter scale at age 10 in the BCS) and smoking at age 42. Furthermore, in recent work Hsieh and van Kippersluis (2018) have shown personality to be a key mechanism through which peers affect smoking behaviour.

Conditional on socio-emotional skills, cognition has limited predictive power for these behaviours, and only for girls.[21] This is in line with the evidence in Conti and Heckman (2010), who show that not accounting for non-cognitive traits (a self-regulation factor measured at age 10) overestimates the importance of cognition for predicting health and health behaviours, using data from the British cohort study. Conti and Hansman (2013) use rich data on child personality and socio-emotional traits collected at ages 7, 11 and 16 in the 1958 British birth

[20] Unfortunately the strength of the association cannot be directly compared, since the outcomes are measured at different ages: 16 and 14 years for BCS and MCS, respectively.

[21] We do not observe significant associations between early socio-emotional skills and other risky behaviours like drug-taking and alcohol consumption. One possible reason might be the relatively young age at which these skills are measured. Results are available upon request.

cohort,[22] and show that these traits rival the importance of cognition in explaining the education gradient in health behaviours (including smoking and BMI). We show that child socio-emotional skills have greater predictive power for health outcomes and behaviours even when measured at an earlier age.

Lastly, for the British Cohort Study, we examine the association between socio-emotional skills at age five and adult education and labour market outcomes. The structure of Table 3.5 is similar to Table 3.4, but it considers educational achievement, employment, and earnings (conditional on being in paid employment) for the BCS cohort members. For these outcomes, the predictive power of cognitive skills outweighs that of socio-emotional skills, whose predictive power diminishes over time (between the ages 34 and 42), and is driven to insignificance after controlling for cognition. This is consistent with the evidence in Conti et al. (2011), who show that cognitive endowments at age 10 are more predictive (than socio-emotional and health ones) for employment and wage outcomes in the BCS. Again, we show that the greater predictive power of cognition for socioeconomic outcomes holds even when considering earlier-life measures of child development. Similarly, Goodman et al. (2015) use data from the same cohort, but exploit a richer set of measures from the age 10 survey. They show that social and emotional skills measured at age 10 are strongly related to a wide range of adult outcomes.

3.7 Conclusion

In this paper we have studied inequality in a dimension of human capital which has received limited attention in the literature so far: socio-emotional skills very early in life. In particular, we have focused on the measurements of these skills at age 5 in two British cohorts born 30 years apart: the one of children born in 1970 (British Cohort Study, BCS) and the one of children born in 2000/1 (Millennium Cohort Study, MCS). We have provided several contributions to the recent but flourishing literature on the determinants and consequences of early human development.

We have taken very seriously the issue of comparability of measurements of socio-emotional skills across cohorts. First, we have selected 11 comparable items across two related scales: the Rutter scale in the BCS, and the Strength and Difficulties Questionnaire (SDQ) in the MCS. After examining the latent structure underlying the items, we have identified by means of exploratory factor analysis two dimensions of socio-emotional skills. We have labeled them ‘internalising’ and ‘externalising’ skills, the former related to the ability of children to focus their concentration and the latter to engage in interpersonal activities.

Second, we have formally tested for measurement invariance of the 11 items across the two externalising and internalising scales, following recent methodological advances in factor analysis with categorical outcomes. We have found only partial support for measurement invariance, with the implication that we have only been able to compare how inequality in these

[22] They use the Rutter scale and the Bristol Social Adjustment Guide.

socio-emotional skills across the two cohorts has changed, but not whether their average level is larger or smaller in one of the two cohorts. These results sound a warning to research in this area which routinely compares levels of skills across different groups (at different times, or of different gender), without first establishing their comparability.

Third, after having computed comparable scores for both externalising and internalising skills, and for both boys and girls, we have compared how inequality in these skills has changed across the 1970 and the 2000 cohort. We have documented for the first time that inequality in these early skills has increased across cohorts, especially for boys. The cross-cohort increase in the gap is more pronounced at the bottom of the distribution (50-10 percentiles). We have also documented changes in conditional skills gaps across cohorts. In particular, the difference in the socio-emotional skills of their children between mothers of higher and lower socio-economic status (education and employment) has increased. The increase in cross-cohort inequality is even starker when comparing children born to mothers who smoked during pregnancy. In both cases, the increase in inequality is particularly pronounced for boys. On the other hand, the skills penalty arising from the lack of a father figure in the household has substantially declined, especially for girls.

Fourth, we have contributed to the literature on the predictive power of socio-emotional skills by showing that even skills measured at a much earlier age than in previous work are significantly associated with outcomes both in adolescence and adulthood. In particular, socio-emotional skills are more significant predictors of health and health behaviours (smoking and BMI), while cognition has greater predictive power for socioeconomic outcomes (education, employment and wages). Our results show the importance of inequalities in the early years development for the accumulation of health and human capital across the life course.

3.8 Tables

Table 3.1: Subscale of comparable items

Itm.	Factor	Cat.	Title	Rutter Wording (BCS 1970)	SDQ Wording (MCS 2000/1)
1	EXT	3	*Restless*	Very restless. Often running about or jumping up and down. Hardly ever still	Restless, overactive, cannot stay still for long
2	EXT	3	*Squirmy/fidgety*	Is squirmy or fidgety	Constantly fidgeting or squirming
3	EXT	3	*Fights/bullies*	Frequently fights other children + Bullies other children	Often fights with other children or bullies them
4	EXT	3	*Distracted*	Cannot settle to anything for more than a few moments	Easily distracted, concentration wanders
5	EXT	2	*Tantrums*	Has temper tantrums	Often has temper tantrums or hot tempers
6	EXT	2	*Disobedient*	Is often disobedient	(+) Generally obedient, usually does what adults request
7	INT	3	*Worried*	Often worried, worries about many things	Many worries, often seems worried
8	INT	3	*Fearful*	Tends to be fearful or afraid of new things or new situations	Nervous or clingy in new situations, easily loses confidence
9	INT	3	*Solitary*	Tends to do things on his/her own, rather solitary	Rather solitary, tends to play alone
10	INT	3	*Unhappy*	Often appears miserable, unhappy, tearful or distressed	Often unhappy, down-hearted or tearful
11	INT	2	*Aches*	Complains of headaches + Complains of stomach-ache or has vomited	Often complains of head- aches, stomach-ache or sickness

Notes: *Itm.* is item number. *Factor* is the latent construct to which the item loads – EXT is Externalising skills, INT is Internalising skills. *Cat.* is the number of categories in which the item is coded – 2 denotes a binary item (applies/does not apply) and 3 denotes a 3-category item. *Title* is a short label for the item. *Wording* columns show the actual wording in the scales used in each of the cohort studies. Items denoted by (+) are positively worded in the original scale.

Table 3.2: Quantile differences in scores

	Males		Females	
	BCS (1970)	MCS (2000/1)	BCS (1970)	MCS (2000/1)
Externalising				
90-10	2.08	2.47	2.09	2.22
75-25	1.08	1.36	1.12	1.23
90-50	1.00	1.14	1.02	0.95
50-10	1.08	1.33	1.07	1.27
Internalising				
90-10	1.71	2.28	1.86	1.86
75-25	0.91	1.12	1.01	0.92
90-50	0.73	0.89	0.82	0.72
50-10	0.97	1.38	1.04	1.14

Notes: The table shows differences between quantiles of the distribution of socioemotional skills, by gender and cohort. The distribution is a factor score obtained from the factor model in Section 3.4. These distributions are shown in Figure 3.1.

Table 3.3: Determinants of Socioemotional Skills

	Externalising						Internalising					
	Males			Females			Males			Females		
	(1) BCS	(2) MCS	(3) p-value	(4) BCS	(5) MCS	(6) p-value	(7) BCS	(8) MCS	(9) p-value	(10) BCS	(11) MCS	(12) p-value
Maternal education (5)												
Post-compulsory	0.089***	0.141***	[0.246]	0.099***	0.147***	[0.270]	0.072***	0.133***	[0.127]	0.048**	0.075**	[0.491]
	(0.026)	(0.041)		(0.027)	(0.037)		(0.022)	(0.039)		(0.025)	(0.032)	
Maternal employment (5)												
Employed	0.018	0.111***	[0.030]	−0.009	0.076**	[0.041]	0.041**	0.148***	[0.005]	0.031	0.127***	[0.011]
	(0.024)	(0.040)		(0.025)	(0.037)		(0.020)	(0.037)		(0.021)	(0.031)	
Father occ. (5) - White collar = 0												
Blue collar	−0.195***	−0.112***	[0.074]	−0.117***	−0.072**	[0.312]	−0.076***	−0.053	[0.588]	−0.101***	−0.021	[0.047]
	(0.027)	(0.043)		(0.027)	(0.039)		(0.023)	(0.040)		(0.025)	(0.034)	
No father figure	−0.280***	−0.225***	[0.512]	−0.380***	−0.179***	[0.010]	−0.200***	−0.193***	[0.926]	−0.244***	−0.123**	[0.083]
	(0.063)	(0.057)		(0.059)	(0.053)		(0.053)	(0.057)		(0.054)	(0.048)	
Maternal background (0)												
Age	0.014***	0.012***	[0.711]	0.014***	0.019***	[0.290]	0.008***	0.011***	[0.430]	0.009***	0.011***	[0.474]
	(0.002)	(0.004)		(0.003)	(0.003)		(0.002)	(0.003)		(0.002)	(0.003)	
Unmarried	0.065	−0.107**	[0.017]	0.025	−0.066*	[0.183]	0.115**	−0.029	[0.025]	0.036	−0.040	[0.221]
	(0.059)	(0.044)		(0.060)	(0.041)		(0.050)	(0.042)		(0.055)	(0.035)	
Nonwhite child	−0.161**	−0.110*	[0.602]	−0.029	−0.131**	[0.238]	0.025	−0.050	[0.391]	0.081	−0.140**	[0.005]
	(0.076)	(0.063)		(0.067)	(0.054)		(0.063)	(0.057)		(0.062)	(0.049)	
Pregnancy												
Firstborn	−0.121***	0.006	[0.009]	−0.070**	0.036	[0.025]	−0.186***	−0.085**	[0.020]	−0.161***	−0.032	[0.002]
	(0.029)	(0.045)		(0.031)	(0.041)		(0.024)	(0.042)		(0.028)	(0.037)	
Mother smoked in pregnancy	−0.144***	−0.229***	[0.092]	−0.110***	−0.156***	[0.354]	−0.077***	−0.176***	[0.028]	−0.036	−0.060	[0.592]
	(0.026)	(0.051)		(0.025)	(0.048)		(0.021)	(0.048)		(0.023)	(0.041)	
(log) Birthweight	0.146**	0.311***	[0.190]	0.186**	0.362***	[0.145]	0.095	0.387***	[0.009]	0.123*	0.078	[0.674]
	(0.073)	(0.113)		(0.078)	(0.102)		(0.061)	(0.109)		(0.070)	(0.088)	
Adj. R^2	0.062	0.081		0.056	0.090		0.042	0.068		0.042	0.052	
Num. obs.	4565	2759		4313	2620		4565	2759		4313	2620	

Notes: The table shows coefficients from linear regressions of children's socioemotional skills at five years of age on family background information. The dependent variable is a factor score obtained from the factor model in Section 3.4. Col. (1) and (2) show coefficients and standard errors in parentheses, for male children in the BCS and MCS cohorts separately. The latter are obtained using 1,000 bootstrap repetitions, taking into account the factor estimation stage that precedes the regression. Col. (3) shows the p-value of a test that the coefficient is the same in the two cohorts. Col. (4) to (6) repeat for female children. Col. (7) to (12) repeat for internalising skills. All estimates additionally control for region of birth, mother height, number of previous stillbirths at child's birth, preterm birth, a dummy for missing gestational age, and number of other children in the household at child age 5. See Table 3.A2 for a description of the variables used.

Table 3.4: Predictors of adolescent outcomes

	Males			Females		
	Mean	Coefficients		Mean	Coefficients	
	(1)	(2)	(3)	(4)	(5)	(6)
Tried smoking (BCS - 16)	.524			.586		
Externalising skills (5)		$-.073^{***}$	$-.081^{***}$		$-.068^{***}$	$-.077^{***}$
		(.023)	(.023)		(.021)	(.022)
Internalising skills (5)		$.055^{**}$	$.060^{**}$		$.039^{*}$	$.045^{*}$
		(.027)	(.028)		(.023)	(.024)
Cognitive skills (5)			.010			.012
			(.019)			(.017)
Adj. R^2		0.032	0.032		0.048	0.046
Observations		1197	1123		1693	1581
BMI (BCS - 16)	20.9			21.2		
Externalising skills (5)		$-.178$	$-.227^{*}$		$-.225^{*}$	$-.222^{*}$
		(.119)	(.124)		(.124)	(.129)
Internalising skills (5)		.036	.062		$.280^{**}$	$.234^{*}$
		(.141)	(.148)		(.138)	(.142)
Cognitive skills (5)			.021			$-.093$
			(.110)			(.104)
Adj. R^2		0.017	0.018		0.023	0.023
Observations		1640	1531		1873	1757
Tried smoking (MCS - 14)	.125			.151		
Externalising skills (5)		$-.043^{***}$	$-.041^{***}$		$-.031^{**}$	$-.029^{**}$
		(.012)	(.012)		(.012)	(.012)
Internalising skills (5)		.017	.018		.009	.012
		(.012)	(.012)		(.014)	(.014)
Cognitive skills (5)			.000			$-.031^{**}$
			(.010)			(.012)
Adj. R^2		0.056	0.054		0.050	0.051
Observations		1959	1936		1986	1982
BMI (MCS - 14)	20.7			21.7		
Externalising skills (5)		$-.327^{**}$	$-.311^{**}$		$-.405^{***}$	$-.382^{***}$
		(.138)	(.139)		(.141)	(.142)
Internalising skills (5)		.064	.091		$.354^{**}$	$.366^{**}$
		(.147)	(.149)		(.157)	(.157)
Cognitive skills (5)			.014			$-.186$
			(.119)			(.152)
Adj. R^2		0.025	0.024		0.045	0.045
Observations		1965	1936		1893	1886

Notes: The table shows coefficients from linear regressions of cohort members' adolescent outcomes on their externalising and internalising socioemotional skills at five years of age. Col. (1) shows the mean of the outcome for males. Col. (2) regresses the outcome on the scores obtained from the factor model in Section 3.4. Col. (3) additionally controls for cognitive ability at age five. This is a simple factor score obtained by aggregating the available cognitive measures. All standard errors in parentheses are obtained using 1,000 bootstrap repetitions, taking into account the factor estimation stage that precedes the regression. Col. (4) to (6) repeat for female cohort members. All estimates additionally control for region of birth, maternal education (5), maternal employment (5), father occupation (5), maternal background (age, height, nonwhite ethnicity, number of children in HH), pregnancy (firstborn child, number of previous stillbirths, mother smoked in pregnancy, preterm birth, (log) birth weight). See Table 3.A2 for a description of the variables used.

Table 3.5: Predictors of adult outcomes – BCS

	Males			Females		
	Mean	Coefficients		Mean	Coefficients	
	(1)	(2)	(3)	(4)	(5)	(6)
Higher education (34)	.430			.426		
Externalising skills (5)		.043**	.024		.068***	.053**
		(.021)	(.022)		(.021)	(.021)
Internalising skills (5)		−.032	−.026		−.017	−.028
		(.027)	(.027)		(.023)	(.024)
Cognitive skills (5)			.089***			.113***
			(.017)			(.017)
Adj. R^2		0.083	0.099		0.101	0.120
Observations		1320	1237		1691	1589
Employed (42)	.932			.828		
Externalising skills (5)		.012	.010		.014	.014
		(.011)	(.011)		(.016)	(.017)
Internalising skills (5)		.022*	.020		.024	.017
		(.014)	(.014)		(.018)	(.019)
Cognitive skills (5)			.023**			.037***
			(.010)			(.014)
Adj. R^2		0.056	0.052		0.010	0.014
Observations		1294	1216		1677	1571
(log) Gross weekly pay (42)	6.474			5.775		
Externalising skills (5)		.047	.047		.009	.003
		(.037)	(.036)		(.042)	(.044)
Internalising skills (5)		−.044	−.082*		.051	.041
		(.044)	(.043)		(.046)	(.047)
Cognitive skills (5)			.064**			.137***
			(.029)			(.033)
Adj. R^2		0.057	0.068		0.046	0.061
Observations		918	865		1198	1122

Notes: The table shows coefficients from linear regressions of BCS cohort members' adult outcomes on their externalising and internalising socioemotional skills at five years of age. Col. (1) shows the mean of the outcome for males. Col. (2) regresses the outcome on the scores obtained from the factor model in Section 3.4. Col. (3) additionally controls for cognitive ability at age five. This is a simple factor score obtained by aggregating the available cognitive measures. All standard errors in parentheses are obtained using 1,000 bootstrap repetitions, taking into account the factor estimation stage that precedes the regression. Col. (4) to (6) repeat for female cohort members. All estimates additionally control for region of birth, maternal education (5), maternal employment (5), father occupation (5), maternal background (age, height, nonwhite ethnicity, number of children in HH), pregnancy (firstborn child, number of previous stillbirths, mother smoked in pregnancy, preterm birth, (log) birth weight). See Table 3.A2 for a description of the variables used.

3.9 Figures

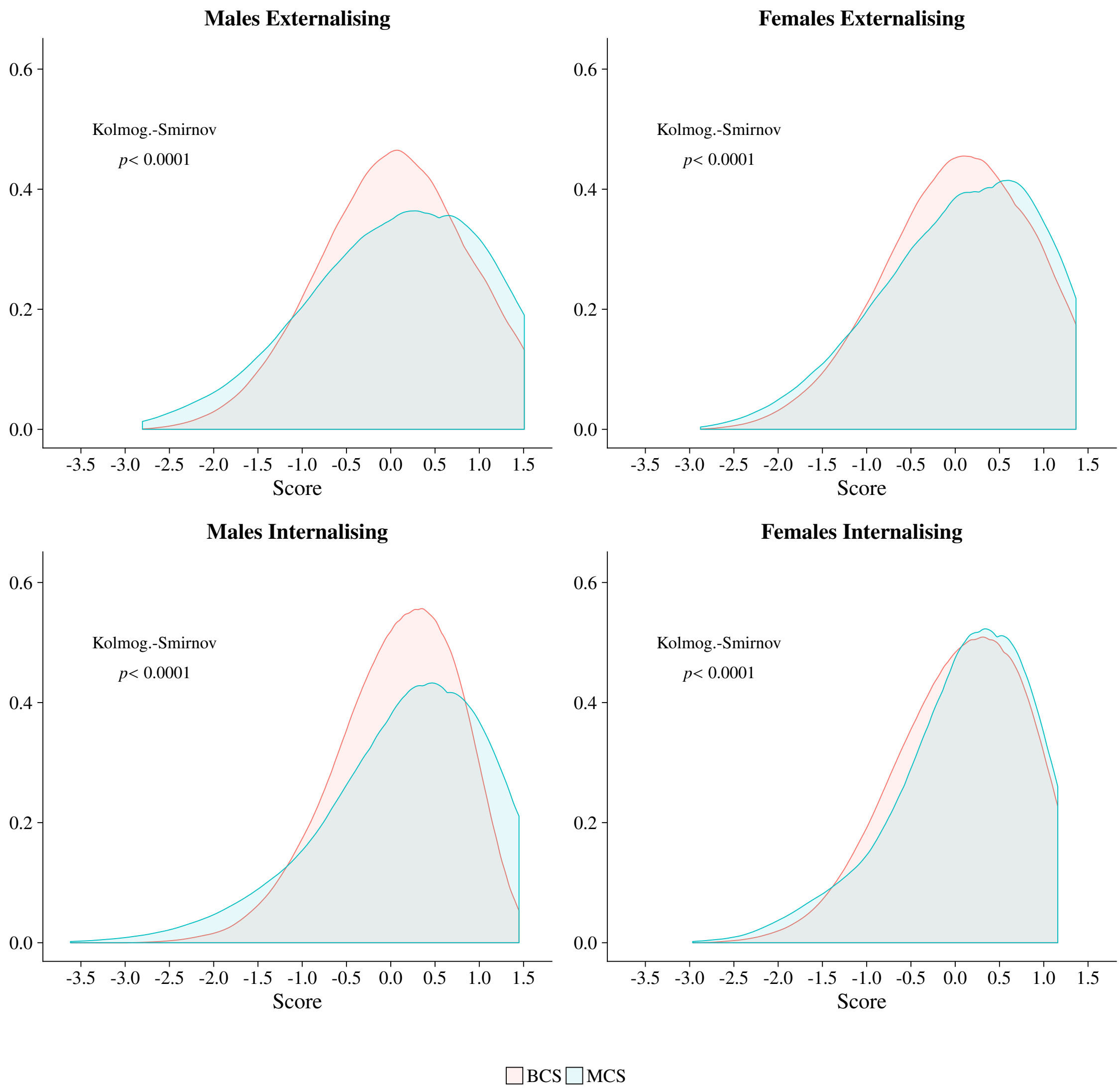

Figure 3.1: Distribution of factor scores

Notes: The figure shows the distribution of the externalising and internalising socioemotional skills scores at age five obtained from the factor model, by gender and cohort. The scores are estimated from parameter estimates in Table 3.A9, using an Empirical Bayes Modal approach. Higher scores correspond to *better* skills. The distribution is estimated nonparametrically, using an Epanechnikov kernel. The figure also reports the p value from Kolmogorov-Smirnov tests of equality between the distribution in BCS and MCS.

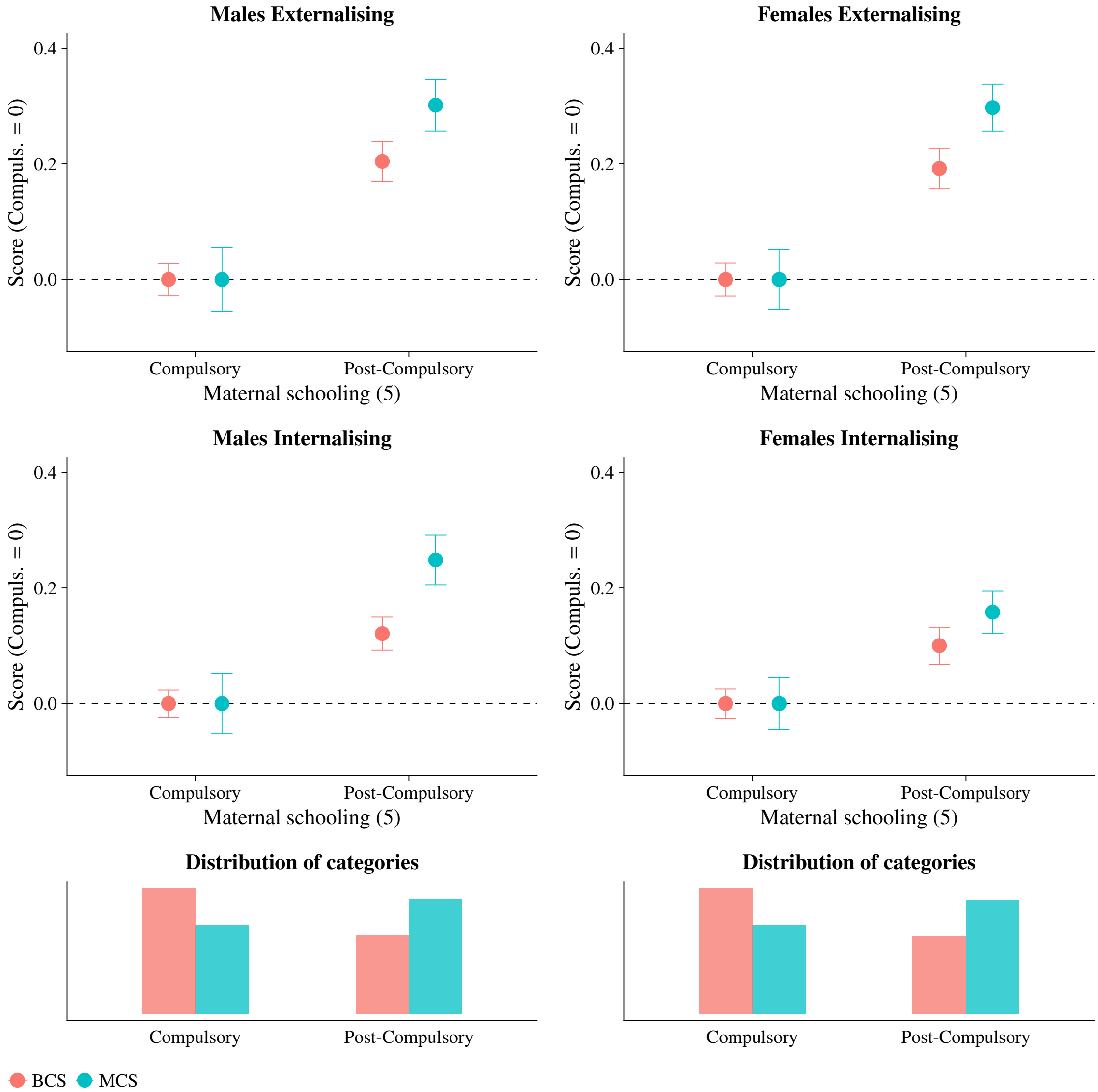

Figure 3.2: Skill inequality by mother's education

Notes: The figure shows unconditional mean values of socioemotional skills scores by gender, cohort, and mother's education at age five. Mother's education is a dummy for whether the mother continued schooling past the minimum leaving age, based on her date of birth. The four panels on top present mean and 95% confidence intervals. Given that we cannot compare means of skills, all scores are normalised to take value zero for the 'Compulsory' category, so that the gradient is emphasised. The bottom two panels present the unconditional distribution of mother's education.

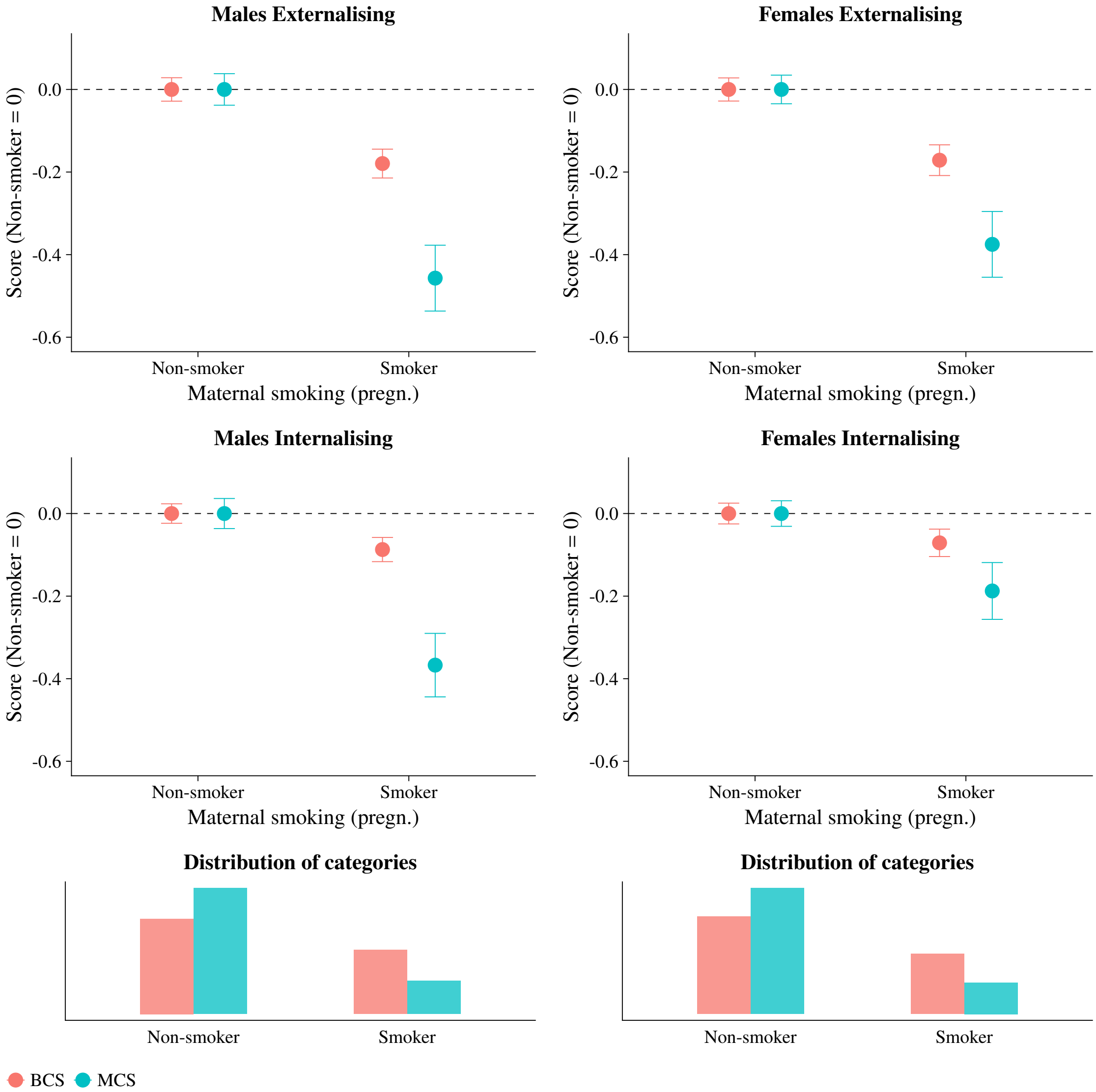

Figure 3.3: Skill inequality by mother's pregnancy smoking

Notes: The figure shows unconditional mean values of socioemotional skills scores by gender, cohort, and mother's pregnancy smoking. Mother's education is a dummy for whether the mother reported smoking during pregnancy. The four panels on top present mean and 95% confidence intervals. Given that we cannot compare means of skills, all scores are normalised to take value zero for the 'Non-smoker' category, so that the gradient is emphasised. The bottom two panels present the unconditional distribution of mother's pregnancy smoking

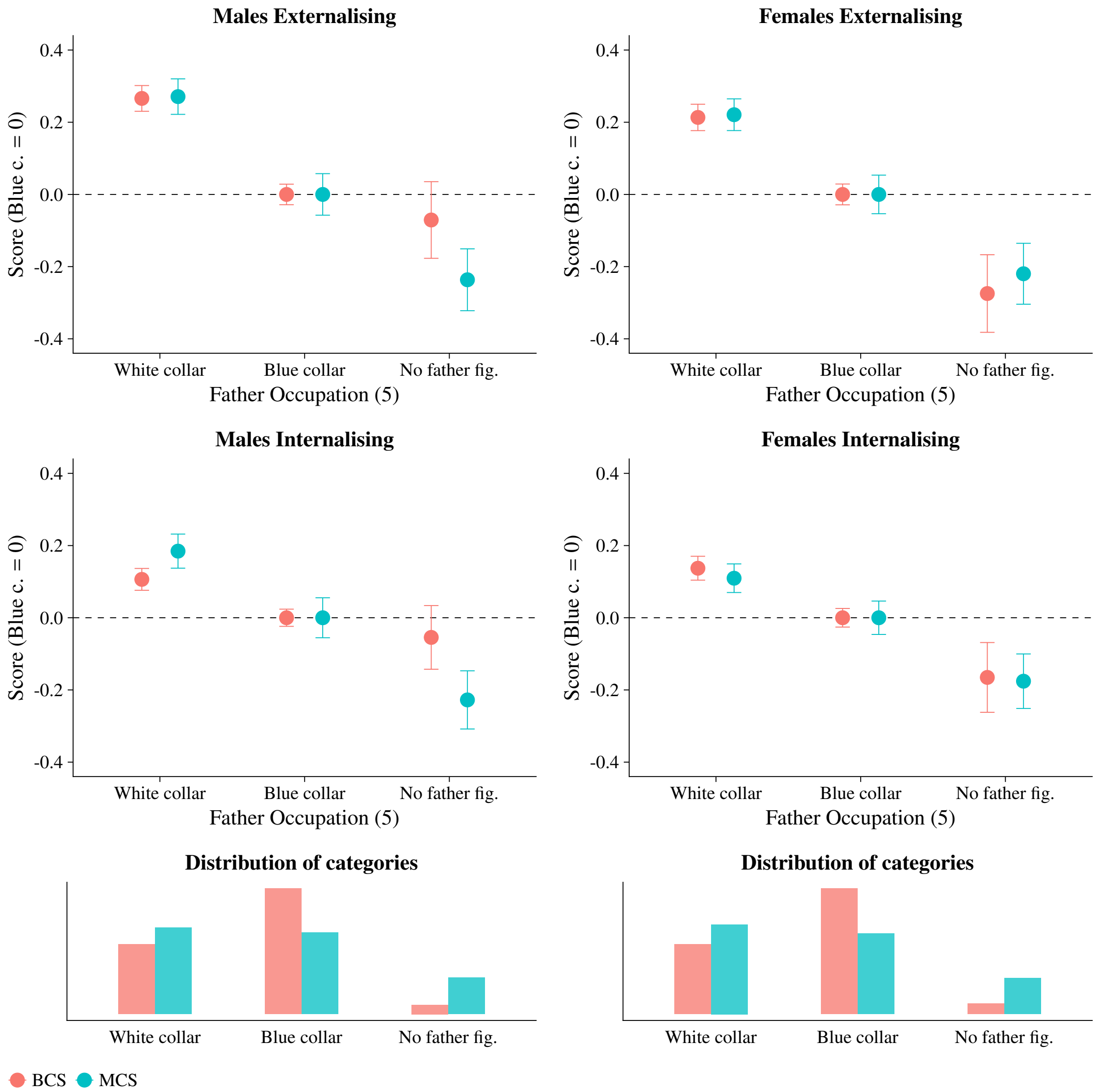

Figure 3.4: Skill inequality by father's occupation

Notes: The figure shows unconditional mean values of socioemotional skills scores by gender, cohort, and father's occupation at age five. Father's occupation is based on Registrar General's social class, with classes I to III Non Manual being 'White collar' and classes III Manual to V (plus 'other') being 'Blue collar'. 'No father figure' is defined as absence of a male figure living in the household. The four panels on top present mean and 95% confidence intervals. Given that we cannot compare means of skills, all scores are normalised to take value zero for the 'Blue collar' category, so that the gradient is emphasised. The bottom two panels present the unconditional distribution of father's occupation.

3.10 Appendix

3.10.1 Deriving a common scale of socioemotional skills

In the BCS data, maternal reports on child socioemotional skills are measured using the Rutter A Scale (Rutter et al., 1970) – see Panel A of Table 3.A1. The Rutter items are rated on three levels: 'Does not apply', 'Somewhat applies', 'Certainly applies'. Since they are all behaviours indicating lower skills, we encode all of them in reverse, i.e. 'Certainly applies' = 0, 'Somewhat applies' = 1, 'Does not apply' = 2. We augment the 19-item Rutter Scale with four additional parent-reported questions from the parental questionnaire, items A to D. These are rated on 4 levels: 'Never in the last 12 months', 'less than once a month', 'at least once a month', 'at least once a week'. we recode these into binary indicators, with 'Never' and 'Less than once a month' to 1 and zero otherwise. To increase comparability between the two scales, we merge together two pairs of items: 4 and 19 (to mirror SDQ item 12 "Often fights with other children or bullies them"), and A and B (to mirror SDQ item 3 "Often complains of head-aches, stomach-ache or sickness"). We assign the lowest category among the two original items to the newly obtained item. We also recode items 5 and 14 to binary instead of three categories. These items are recorded with a positive phrasing in SDQ, so a 3-category split would be harder to compare.

In MCS, we use the 25-item strengths and difficulties questionnaire (Goodman, 1997) – see Panel B of Table 3.A1. All items are recorded on a 4-point scale: 'Not true', 'Somewhat true', 'Certainly true', 'Can't say'. We set the latter option to missing and recode the rest in ascending order of skill as for the BCS items, i.e. 'Certainly true' = 0, 'Somewhat true' = 1, 'Not true' = 2. For comparability with the BCS Rutter scale, we dichotomise items 3 and 5 to make them comparable with , and dichotomise and invert items 7, and 14.

3.10.2 Measurement invariance details

Alternative parameterisations for the configural model

There are infinite ways to parameterise the configural model defined by (3.4.1) and (3.4.2). Widely used parameterisations are:

◇ Delta parameterisation [WEΔ] (Wu and Estabrook, 2016)
For all groups:

$$\text{diag}(\boldsymbol{\Phi}) = \boldsymbol{I}, \quad \boldsymbol{\kappa} = \boldsymbol{0}, \quad \boldsymbol{\nu} = \boldsymbol{0}, \quad \text{and diag}(\boldsymbol{\Sigma}) = \boldsymbol{I}.$$

◇ Theta parameterisation [WEΘ] (Wu and Estabrook, 2016)
For all groups:

$$\text{diag}(\boldsymbol{\Phi}) = \boldsymbol{I}, \quad \boldsymbol{\kappa} = \boldsymbol{0}, \quad \boldsymbol{\nu} = \boldsymbol{0}, \quad \text{and diag}(\boldsymbol{\Psi}) = \boldsymbol{I}.$$

◇ Anchored parameterisation [MT] (Millsap and Yun-Tein, 2004)

- For all groups, normalise a reference loading to 1 for each factor

- Set invariant across groups one threshold per item (e.g. $\tau_{0,Ai} = \tau_{0,Bi}$), and an additional threshold in the reference items above
- In the first group: $\boldsymbol{\kappa_A} = \mathbf{0}$, diag$(\boldsymbol{\Sigma}_A) = \boldsymbol{I}$
- Set all intercepts $\boldsymbol{\nu}$ to zero

The first two parameterisations (WEΔ and WEΘ) normalise the mean and variance of factors to the same constants in both groups, and they leave all loadings and intercepts to be freely estimated; they only differ in whether the additional required normalisation is imposed on the variances of the error terms ($\boldsymbol{\Psi}$) or on the diagonal of the covariance matrix of the measures ($\boldsymbol{\Sigma}$). The MT parameterisation instead proceeds by identifying parameters in one group first, and then imposing cross-group equality constraints to identify parameters in other groups (Wu and Estabrook, 2016).

Identification of models with different levels of invariance

In the case where available measures are continuous, MI analysis is straightforward (van de Schoot et al., 2012). The hierarchy of the nested models usually proceeds by testing loadings first, and then intercepts (to establish *metric* and *scalar* invariance – see Vandenberg and Lance, 2000).

Invariance of systems with categorical measures, such as the scale we examine in this paper, is less well understood. In particular, the lack of explicit location and scale in the measures introduces an additional set of parameters compared to the continuous case (thresholds τ). This makes identification reliant on more stringent normalisations. A first comprehensive approach for categorical measures was proposed by Millsap and Yun-Tein (2004). New identification results in Wu and Estabrook (2016) indicate that, in the categorical case, invariance properties cannot be examined by simply restricting one set of parameters at a time. This is because the identification conditions used in the configural baseline model, while being minimally restrictive on their own, become binding once certain additional restrictions are imposed. In light of this, they propose models that identify structures of different invariance levels. They find that some restrictions cannot be tested alone against the configural model, because the models they generate are statistically equivalent. This is true of loading invariance, and also of threshold invariance in the case when the number of categories of each ordinal item is 3 or less. Furthermore, they suggest that comparison of both latent means and variances requires invariance in loadings, thresholds, and intercepts. A summary of the approach by Wu and Estabrook (2016) is available in Table 3.A3.

3.10.3 Appendix tables

Table 3.A1: Behavioural screening scales in the BCS and MCS five-year surveys

Panel A: Rutter A Scale (Rutter et al., 1970) – British Cohort Study (1975) five-year survey

1. Very restless. Often running about or jumping up and down. Hardly ever still.†
2. Is squirmy or fidgety.†
3. Often destroys own or others' belongings.
4. Frequently fights other children.†
5. Not much liked by other children.
6. Often worried, worries about many things.†
7. Tends to do things on his/her own, is rather solitary.†
8. Irritable. Is quick to fly off the handle.
9. Often appears miserable, unhappy, tearful or distressed.†
10. Sometimes takes things belonging to others.
11. Has twitches, mannerisms or tics of the face or body.
12. Frequently sucks thumb or finger.
13. Frequently bites nails or fingers.
14. Is often disobedient.†
15. Cannot settle to anything for more than a few moments.†
16. Tends to be fearful or afraid of new things or new situations.†
17. Is over fussy or over particular.
18. Often tells lies.
19. Bullies other children.†

A. Complains of headaches.†
B. Complains of stomach-ache or has vomited.†
C. Complains of biliousness.
D. Has temper tantrums (that is, complete loss of temper with shouting, angry movements, etc.).†

Panel B: Strength and Difficulties Questionnaire (Goodman, 1997) – Millennium Cohort Study (2000/1) five-year survey

1. Considerate of other people's feelings.
2. Restless, overactive, cannot stay still for long.†
3. Often complains of head- aches, stomach-ache or sickness.†
4. Shares readily with other children (treats, toys, pencils, etc.).+
5. Often has temper tantrums or hot tempers.†
6. Rather solitary, tends to play alone.†
7. Generally obedient, usually does what adults request.†+
8. Many worries, often seems worried.†
9. Helpful if someone is hurt, upset or feeling ill.+
10. Constantly fidgeting or squirming.†
11. Has at least one good friend.+
12. Often fights with other children or bullies them.†
13. Often unhappy, down-hearted or tearful.†
14. Generally liked by other children.+
15. Easily distracted, concentration wanders.†
16. Nervous or clingy in new situations, easily loses confidence.†
17. Kind to younger children.+
18. Often lies or cheats.
19. Picked on or bullied by other children.
20. Often volunteers to help others (parents, teachers, other children).+
21. Thinks things out before acting.+
22. Steals from home, school or elsewhere.
23. Gets on better with adults than with other children.
24. Many fears, easily scared.
25. Sees tasks through to the end, good attention span.+

Notes: Items denoted by + are positively worded in the original scale. Items denoted by † are retained in the new comparable scale.

Table 3.A2: Description of harmonised variables

Variable Group	Age	Variable	Note
Maternal education	5	Post-compulsory schooling[d]	Whether mother continued schooling past the compulsory age, based on her year of birth. School leaving age in England was changed from 14 to 15 in 1947 and from 15 to 16 in 1972.
Maternal employment	5	Employed[d]	Includes full time and part time
Father occupation	5	White collar (I-IIINM)[d] Blue collar (IIIM-V-other)[d] No father figure[d]	Based on father's Registrar General Social Class classification of occupations. White collar includes I (Professional), II (Managerial/technical), IIINM (Skilled non-manual). Blue collar includes IIIM (Skilled manual), IV (Partly skilled), V (Unskilled), Other, Unemployed, and Armed forces. No father figure is a dummy for children whose father does not live in the same household. Father's social class was recorded using the SOC2000 classification in MCS. We use the derivation matrices kindly provided by David Pevalin at ISER (available at https://www.iser.essex.ac.uk/archives/nssec/derivations-of-social-class) to map SOC2000 into Registrar General Social Class.
Maternal background	0/5	Mother's age at birth Mother's height (cm) Mother unmarried at birth[d] Child nonwhite ethnicity[d] Number of children in HH	All variables are self-reported by the mother at birth, except for number of children in household (at five years old). Unmarried is only based on marital status, and includes cohabitation.
Pregnancy	0	Child is firstborn[d] Number of previous stillbirths Mother smoked in pregnancy[d] Preterm birth (under 37 weeks gestation)[d] (log) birth weight (kg)	Parity, stillbirths, and smoking are self-reported by the mother. Gestational length and birth weight are from hospital records.
Cognitive skills	5		Based on test batteries administered to the cohort member at five. Three tests are used for BCS children: Copy Designs child is asked to copy simple designs adjacently), Human Figure Drawing (child draws an entire human figure), English Picture Vocabulary Test (child identifies the picture referring to a word among four pictures). Three *different* tests are used in the MCS: BAS Naming Vocabulary (child is shown a series of pictures and asked to name it), BAS Picture Similarity (child is shown a row of 4 pictures on a page and places a card with a fifth picture under the one most similar to it), BAS Pattern Construction (child constructs a design by putting together flat squares or solid cubes with patterns on each side).
Adolescent outcomes	16 (BCS) 14 (MCS)	Child tried smoking[d] Body Mass Index (BMI)	Smoking is self reported by the child. Height and weight are taken as part of a medical examination.
Adult outcomes (BCS only)	34 42 34, 42	Higher education[d] Employed[d] (log) gross weekly pay	Higher education is defined on having a university degree or its vocational equivalent. It corresponds to level 4 or 5 in the National Vocational Qualification (NVQ) equivalence. Employed is a dummy for being in paid employment or self-employment, either full or part time. Gross weekly pay is weekly pre-tax pay from the respondent's main activity, conditional on being a paid employee.

Notes: Variables denoted by [d] are binary or categorical.

Table 3.A3: Parameterisations for measurement invariance

Invariance level	Description	Restrictions
Configural (WEΘ)	· Minimally restrictive model for identification	For all groups: $\text{diag}(\boldsymbol{\Phi}) = \boldsymbol{I}$; $\boldsymbol{\kappa} = \boldsymbol{0}$; $\boldsymbol{\nu} = \boldsymbol{0}$; $\text{diag}(\boldsymbol{\Psi}) = \boldsymbol{I}$
Threshold invariance	· Restricts thresholds τ to be equal across groups · Statistically equivalent to configural (when measures have 3 categories or less)	$\tau_{1,ci} = \tau_{1,c'i}$ for all items, $\forall c, c'$ $\tau_{2,ci} = \tau_{2,c'i}$ for non-binary items, $\forall c, c'$ For all groups: $\text{diag}(\boldsymbol{\Phi}) = \boldsymbol{I}$; $\boldsymbol{\kappa} = \boldsymbol{0}$ For ref. group A: $\boldsymbol{\nu}_A = \boldsymbol{0}$ $\text{diag}(\boldsymbol{\Sigma}_A) = \boldsymbol{I}$
Threshold and Loading invariance	· Restricts thresholds τ and loadings λ to be equal across groups · Allows comparison of latent factor variances	$\tau_{1,ci} = \tau_{1,c'i}$ for all items, $\forall c, c'$ $\tau_{2,ci} = \tau_{2,c'i}$ for non-binary items, $\forall c, c'$ $\lambda_{ci} = \lambda_{c'i}$ for all items, $\forall c, c'$ For all groups: $\boldsymbol{\kappa} = \boldsymbol{0}$ For ref. group A: $\boldsymbol{\nu}_A = \boldsymbol{0}$ $\text{diag}(\boldsymbol{\Sigma}_A) = \boldsymbol{I}$ $\text{diag}(\boldsymbol{\Phi}_A) = \boldsymbol{I}$
Threshold, Loading, and Intercept invariance	· Restricts thresholds τ and loadings λ to be equal across groups · Restricts intercepts ν to zero in both groups · Allows comparison of latent factor variances *and* means	$\tau_{1,ci} = \tau_{1,c'i}$ for all items, $\forall c, c'$ $\tau_{2,ci} = \tau_{2,c'i}$ for non-binary items, $\forall c, c'$ $\lambda_{ci} = \lambda_{c'i}$ for all items, $\forall c, c'$ For all groups: $\boldsymbol{\nu} = \boldsymbol{0}$ For ref. group A: $\boldsymbol{\kappa}_A = \boldsymbol{0}$ $\text{diag}(\boldsymbol{\Sigma}_A) = \boldsymbol{I}$ $\text{diag}(\boldsymbol{\Phi}_A) = \boldsymbol{I}$

Notes: Adapted from Wu and Estabrook (2016).

Table 3.A4: Item prevalence, by cohort and gender

				Males						Females					
				BCS			MCS			BCS			MCS		
				Cert. Appl.	Smtm. Appl.	Appl.	Cert. True	Smwt. True	True	Cert. Appl.	Smtm. Appl.	Appl.	Cert. True	Smwt. True	True
Itm.	Factor	Cat.	Title	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1	EXT	3	Restless	32.0	40.4		17.3	29.0		25.0	40.4		13.1	24.2	
2	EXT	3	Squirmy/fidgety	12.3	31.8		11.3	29.4		11.3	32.1		9.1	25.4	
3	EXT	3	Fights/bullies	6.6	39.3		1.7	9.1		3.1	28.1		0.9	4.9	
4	EXT	3	Distracted	8.0	30.1		16.0	44.1		6.1	25.6		9.7	38.7	
5	EXT	2	Tantrums			26.3			51.1			19.6			47.1
6	EXT	2	Disobedient			73.7			48.9			64.9			41.7
7	INT	3	Worried	5.6	29.4		2.4	11.8		5.8	31.3		1.5	11.9	
8	INT	3	Fearful	7.0	29.2		11.1	34.3		6.6	30.0		9.8	38.0	
9	INT	3	Solitary	9.6	37.4		6.4	26.1		8.5	35.3		5.0	24.3	
10	INT	3	Unhappy	2.3	18.3		1.6	9.0		3.0	22.5		1.6	8.3	
11	INT	2	Aches			13.3			17.0			14.8			22.2

Notes: The table shows the prevalence by gender and cohort for each item of our novel subscale. *Itm.* is item number. *Factor* is the latent construct to which the item loads – EXT is Externalising skills, INT is Internalising skills. *Cat.* is the number of categories in which the item is coded – 2 denotes a binary item (applies/does not apply) and 3 denotes a 3-category item. *Title* is a short label for the item. *Cert. / Smtm. Appl.* = Certainly / sometimes applies. *Cert. / Smwt. True* = Certainly / somewhat true.

Table 3.A5: Summary statistics

	Males		Females	
	BCS	MCS	BCS	MCS
Mother age (0)	25.935	29.469	25.902	29.423
	(5.413)	(5.564)	(5.277)	(5.658)
Mother height (m)	1.613	1.645	1.614	1.645
	(0.063)	(0.068)	(0.065)	(0.070)
Unmarried (0)	0.049	0.363	0.055	0.353
	(0.216)	(0.481)	(0.227)	(0.478)
Nonwhite child	0.027	0.100	0.035	0.101
	(0.163)	(0.301)	(0.185)	(0.301)
Number of children in HH (5)	1.560	1.352	1.541	1.309
	(1.138)	(0.995)	(1.125)	(0.967)
Firstborn child	0.369	0.412	0.385	0.423
	(0.483)	(0.492)	(0.487)	(0.494)
Number previous stillbirths	0.023	0.010	0.021	0.011
	(0.156)	(0.098)	(0.147)	(0.113)
Mother smoked in pregnancy	0.401	0.209	0.382	0.200
	(0.490)	(0.407)	(0.486)	(0.400)
Preterm birth	0.040	0.077	0.032	0.068
	(0.197)	(0.267)	(0.175)	(0.252)
Missing gest. age	0.191	0.008	0.180	0.008
	(0.393)	(0.091)	(0.384)	(0.089)
Birthweight (kg)	3.369	3.443	3.254	3.317
	(0.544)	(0.587)	(0.509)	(0.568)
Mother has post-compulsory education (5)	0.386	0.563	0.381	0.560
	(0.487)	(0.496)	(0.486)	(0.496)
Mother is employed (5)	0.426	0.618	0.413	0.614
	(0.495)	(0.486)	(0.492)	(0.487)
Father occupation: blue collar	0.614	0.398	0.610	0.391
	(0.487)	(0.490)	(0.488)	(0.488)
No father figure	0.046	0.179	0.051	0.174
	(0.209)	(0.384)	(0.220)	(0.379)

Notes: The table shows mean (standard deviation) of harmonised variables used in the analysis.

Table 3.A6: Suggested number of factors to retain

	BCS (1970)			MCS (2000/1)		
Approach	All	Males	Females	All	Males	Females
Optimal Coordinates	3	3	3	2	2	2
Acceleration Factor	1	1	1	1	1	1
Parallel Analysis	3	3	3	2	2	2
Kaiser	3	3	3	2	2	2
VSS Compl. 1	2	2	1	1	1	1
VSS Compl. 2	2	2	2	2	2	2
Velicer MAP	1	1	1	2	2	2

Notes: The table compares the optimal number of factors suggested by different aprroaches for our novel scale; scree test based approaches (optimal coordinates, acceleration factor – Raîche et al., 2013), parallel analysis (Horn, 1965), Kaiser's eigenvalue rule (Kaiser, 1960), Very Simple Structure (VSS, Revelle and Rocklin, 1979), Velicer Minimum Average Partial test (MAP, Velicer, 1976).

Table 3.A7: Loadings from exploratory factor analysis

		BCS (1970)		MCS (2000/1)	
Item	Title	Factor 1 (EXT)	Factor 2 (INT)	Factor 1 (EXT)	Factor 2 (INT)
1	Restless	**0.79**	-0.113	**0.927**	-0.077
2	Squirmy/fidgety	**0.67**	0.021	**0.751**	0.036
3	Fights/bullies	**0.499**	0.046	**0.507**	0.237
4	Distracted	**0.629**	0.05	**0.66**	0.037
5	Tantrums	**0.484**	0.177	**0.484**	0.217
6	Disobedient	**0.598**	0.066	**0.563**	-0.011
7	Worried	-0.037	**0.729**	-0.079	**0.784**
8	Fearful	-0.064	**0.595**	-0.03	**0.517**
9	Solitary	0.075	**0.312**	0.002	**0.463**
10	Unhappy	0.249	**0.507**	0.115	**0.741**
11	Aches	0.135	**0.268**	0.033	**0.457**

Notes: The table displays the factor loadings obtained from exploratory factor analysis (EFA) on our novel scale, separately by cohort. The EFA is based on the decomposition of the polychoric correlation matrix, and uses weighted least squares and oblimin rotation.

Table 3.A8: Measurement invariance fit comparison

		Absolute fit						**Relative fit**					
Model	Num. par.	χ^2	RMSE	SRMR	MFI	CFI	G-hat	χ^2 p	ΔRMSE	ΔSRMR	ΔMFI	ΔCFI	ΔG-hat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
A: Entire sample													
Configural	124	1887.3	0.0516	0.0647	0.9443	0.9361	0.9796						
Threshold + Loading Inv	97	2217.7	0.0520	0.0679	0.9348	0.9310	0.9761	0.0000	0.0004	0.0033	-0.0095	-0.0051	-0.0035
Thr. + Load. + Intercept Inv	70	7198.5	0.0908	0.0746	0.7923	0.7693	0.9220	0.0000	0.0392	0.0099	-0.1520	-0.1668	-0.0576
B: 59-61 months sample													
Configural	124	1551.5	0.0527	0.0656	0.9420	0.9285	0.9788						
Threshold + Loading Inv	97	1753.9	0.0520	0.0684	0.9349	0.9268	0.9761	0.0000	-0.0007	0.0028	-0.0071	-0.0017	-0.0026
Thr. + Load. + Intercept Inv	70	4640.1	0.0822	0.0738	0.8261	0.8004	0.9351	0.0000	0.0295	0.0082	-0.1160	-0.1280	-0.0437
C: Males only													
Configural	62	987.1	0.0522	0.0650	0.9432	0.9388	0.9792						
Threshold + Loading Inv	53	1118.1	0.0529	0.0670	0.9357	0.9328	0.9764	0.0000	0.0007	0.0020	-0.0074	-0.0060	-0.0028
Thr. + Load. + Intercept Inv	44	3673	0.0944	0.0731	0.7931	0.7708	0.9223	0.0000	0.0423	0.0081	-0.1500	-0.1681	-0.0569
D: Females only													
Configural	62	900.2	0.0510	0.0644	0.9456	0.9324	0.9801						
Threshold + Loading Inv	53	1087.6	0.0536	0.0686	0.9341	0.9211	0.9758	0.0000	0.0026	0.0043	-0.0115	-0.0113	-0.0043
Thr. + Load. + Intercept Inv	44	3347.5	0.0926	0.0749	0.8003	0.7488	0.9251	0.0000	0.0416	0.0105	-0.1453	-0.1835	-0.0550

Notes: The table presents fit indices for models of different invariance levels, following Wu and Estabrook (2016) Col. (1) displays the number of estimated parameters for each model. Col. (2) and (8) present the value of the χ^2 statistic and the pvalue of the test of equality with respect to the configural model. Col. (3)-(7) and (9)-(13) present alternative fit indices (AFIs), in absolute values and differences from the configural model respectively. RMSE = Root mean squared error of approximation; SRMR = standardised root mean residual; MFI = McDonald non-centrality index; CFI = comparative fit index; G-hat = gamma-hat. Panel A shows results for the whole sample of children in the BCS and MCS cohorts. Panel B is restricted to a subsample of children in the age range of maximum overlap between the two cohorts (59-61 months). Panel C and D are limited to the samples of male and female children respectively.

Table 3.A9: Parameter estimates from factor model with threshold and loading invariance

Panel A: Measurement parameters

		Loadings	**Thresholds**		**Intercepts** (BCS M = 0)			**Variances** (BCS M = 1)		
		λ	τ_1	τ_2	ν			diag(Ψ)		
Item	Factor	All	All	All	BCS F	MCS M	MCS F	BCS F	MCS M	MCS F
X1	EXT	1.218	-0.716	0.894	0.329	1.154	1.503	1.141	0.690	0.902
X2	EXT	1.011	-1.641	-0.215	0.014	0.201	0.394	0.872	0.718	0.785
X3	EXT	0.637	-1.725	-0.156	0.486	1.345	1.951	1.116	0.977	1.248
X4	EXT	0.781	-1.843	-0.375	0.181	-0.556	-0.190	0.880	0.785	0.801
X5	EXT	0.665	-0.787		0.260	-0.730	-0.597	1.000	1.000	1.000
X6	EXT	0.683	0.746		0.303	0.830	1.112	1.000	1.000	1.000
X7	INT	0.759	-1.995	-0.501	-0.101	1.210	0.969	0.858	1.298	0.845
X8	INT	0.511	-1.716	-0.367	0.013	-0.131	-0.282	0.981	1.192	0.890
X9	INT	0.384	-1.400	-0.104	0.059	0.478	0.585	0.941	0.909	0.973
X10	INT	1.135	-3.034	-1.257	-0.207	1.149	1.241	1.030	0.850	1.256
X11	INT	0.420	-1.247		-0.094	-0.021	-0.329	1.000	1.000	1.000

Panel B: Latent variable parameters

	Mean		**Covariance**				Correlation	
	κ		Φ					
	BCS	MCS	BCS		MCS		BCS	MCS
Males								
θ^{EXT}	0.000	0.000	1.000		1.334			
θ^{INT}	0.000	0.000	0.420	1.000	0.878	1.934	0.420	0.546
Females								
θ^{EXT}	0.000	0.000	0.985		1.199			
θ^{INT}	0.000	0.000	0.478	1.012	0.607	1.418	0.478	0.465

Notes: The table presents estimates for the factor model with loadings λ and thresholds τ restricted to be equal across cohorts. Panel A shows estimates of the measurement parameters. Loadings and thresholds are the same across all cohorts. Intercepts are restricted to zero in the reference group, i.e. males in BCS (not shown). Variances of the error terms are restricted to one in the reference group, i.e. males in BCS (not shown), and for the items that only have two categories (5, 6, 11). Panel B shows estimates of the latent variable parameters. Means are restricted to zero in all cohort-gender groups, while variances are restricted to one only in the reference group, i.e. males in BCS.

Table 3.A10: Predictors of adult behaviours, BCS

	Males				Females			
	Mean	Coefficients			Mean	Coefficients		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Daily smoker (42)	.180				.147			
Externalising skills (5)		$-.062^{***}$	$-.059^{***}$			$-.050^{***}$	$-.049^{***}$	
		(.017)	(.018)			(.015)	(.016)	
Internalising skills (5)		.027	.025			$.043^{**}$	$.048^{***}$	
		(.020)	(.021)			(.017)	(.017)	
Externalising (sum score)				$-.040^{***}$				$-.027^{***}$
				(.012)				(.010)
Internalising (sum score)				.004				$.023^{***}$
				(.011)				(.009)
Cognitive skills (5)			$-.022$	$-.022$			$-.032^{***}$	$-.032^{***}$
			(.015)	(.014)			(.012)	(.012)
Adj. R^2		0.044	0.045	0.045		0.037	0.041	0.041
Observations		1294	1216	1216		1678	1572	1572
BMI (42)	27.5				26.1			
Externalising skills (5)		$-.267$	$-.138$			$-.386$	$-.242$	
		(.221)	(.229)			(.261)	(.269)	
Internalising skills (5)		.400	.316			.102	$-.035$	
		(.266)	(.272)			(.289)	(.300)	
Externalising (sum score)				$-.041$				$-.204$
				(.159)				(.176)
Internalising (sum score)				.129				$-.038$
				(.149)				(.153)
Cognitive skills (5)			$-.235$	$-.235$			$-.729^{***}$	$-.728^{***}$
			(.194)	(.192)			(.223)	(.214)
Adj. R^2		0.028	0.024	0.024		0.034	0.047	0.047
Observations		1149	1078	1078		1399	1317	1317

Notes: The table shows coefficients from linear regressions of cohort members' adolescent and adult outcomes on their externalising and internalising socioemotional skills at five years of age. Col. (1) shows the mean of the outcome for males. Col. (2) regresses the outcome on the scores obtained from the factor model in Section 3.4. Col. (3) additionally controls for cognitive ability at age five. This is a simple factor score obtained by aggregating the available cognitive measures. Col. (4) uses sum scores (see Figure 3.1) instead of factor scores. All standard errors in parentheses are obtained using 1,000 bootstrap repetitions, taking into account the factor estimation stage that precedes the regression. Col. (5) to (8) repeat for female cohort members. All estimates additionally control for region of birth, maternal education (5), maternal employment (5), father occupation (5), maternal background (age, height, nonwhite ethnicity, number of children in HH), pregnancy (firstborn child, number of previous stillbirths, mother smoked in pregnancy, preterm birth, (log) birth weight). See Table 3.A2 for a description of the variables used.

3.10.4 Appendix figures

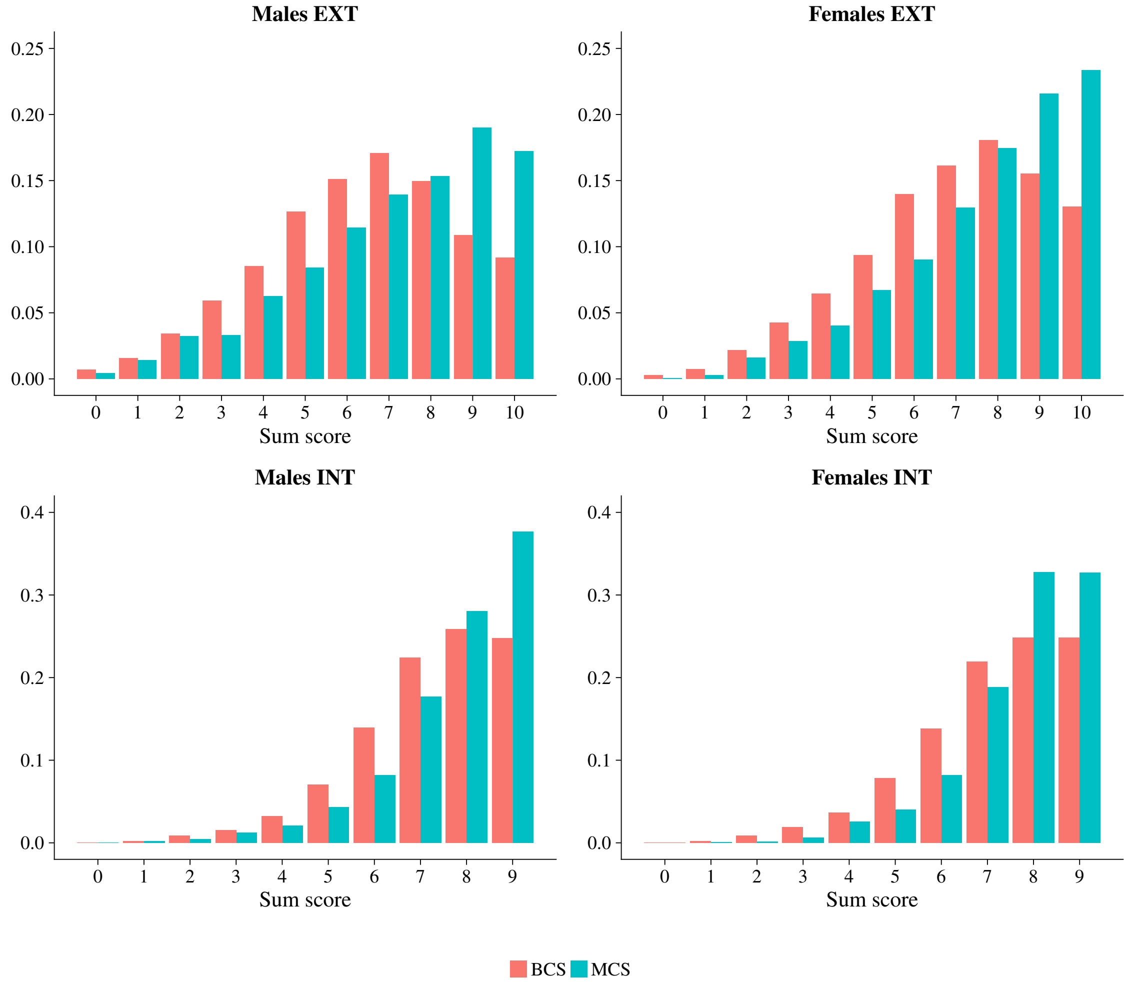

Figure 3.1: Distribution of sum scores

Notes: The figure shows the distribution of the externalising and internalising sum scores at age five, by gender and cohort. The scores are obtained by assigning 0, 1, or 2 points for each item in the scale in Table 3.1. Zero points are assigned for ‘Certainly Applies / True’ responses, one point for ‘Sometimes applies / somewhat true’, and two points for ‘Doesn’t apply’. Only 0 or 1 points are assigned for items that are coded as having two categories (5,6, and 11). Higher scores correspond to *better* skills.

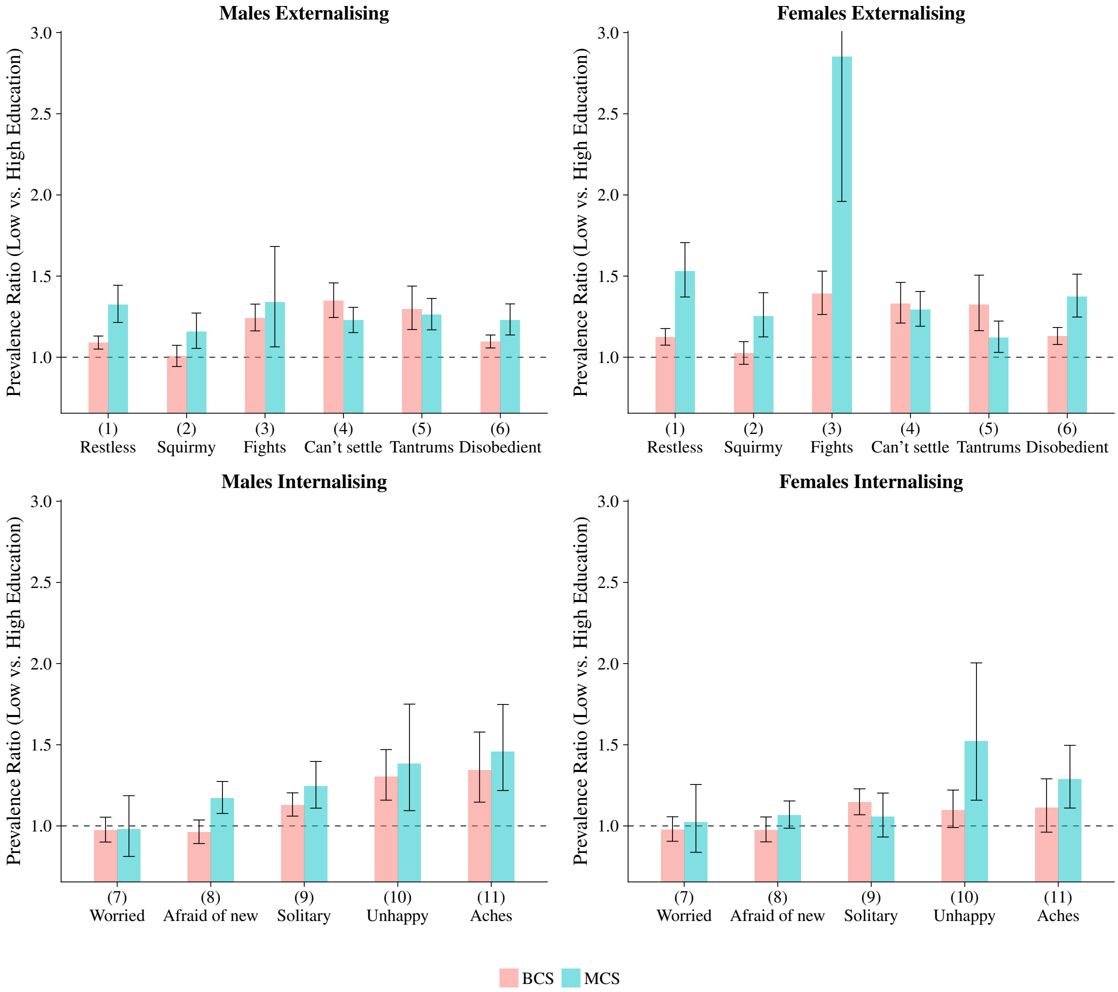

Figure 3.2: Item-level inequality by mother's education

Notes: The graph displays the ratio between the prevalence of each item in our scale in children of educated vs uneducated mothers, by cohort and gender. All items that have three categories in the scale have been dichotomised. For example, if the prevalence of the 'Restless' behaviours among children of mothers with compulsory schooling in the BCS cohort is 7.5%, and 5% among mothers with post-compulsory schooling, the ratio will be 1.5. The error bars represent 95% confidence intervals.

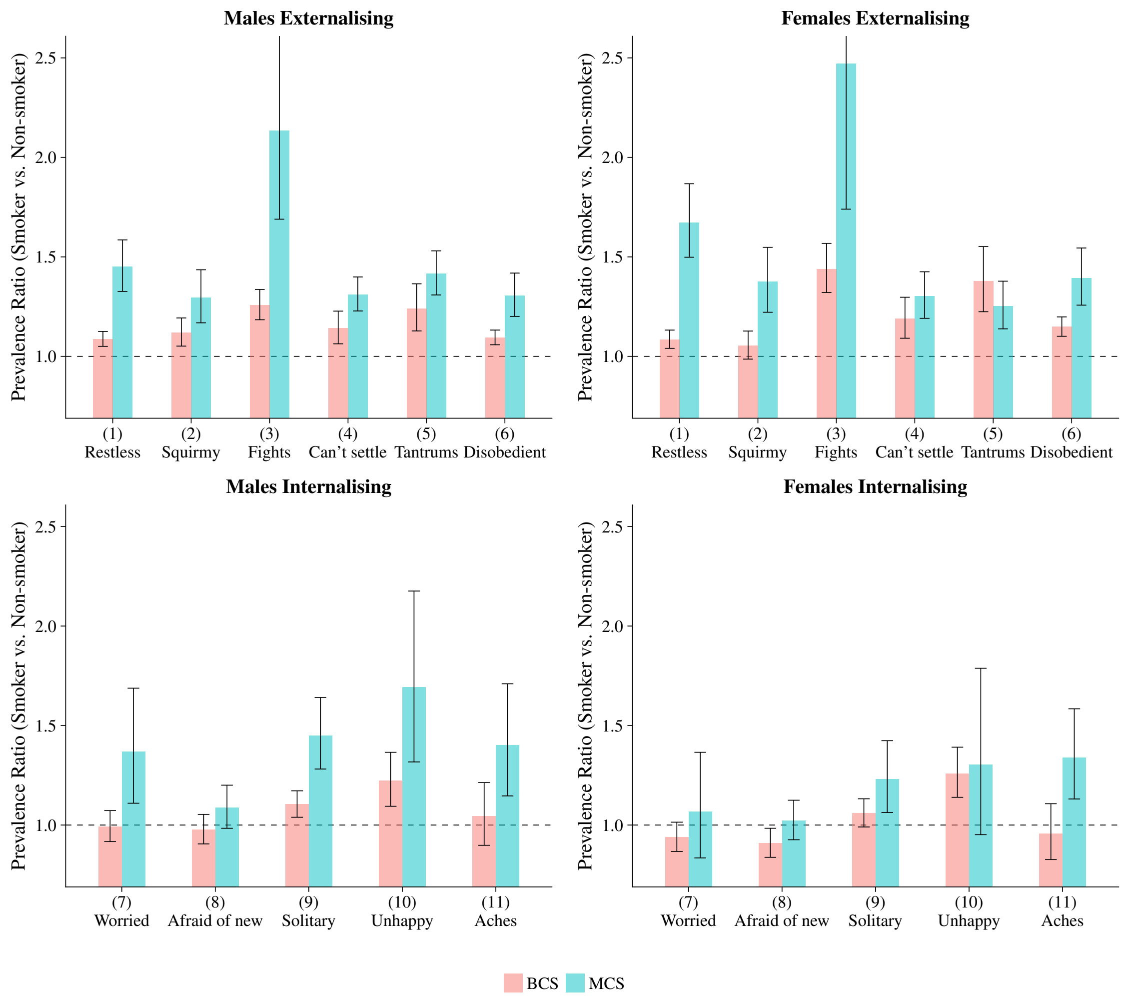

Figure 3.3: Item-level inequality by mother's pregnancy smoking

Notes: The graph displays the ratio between the prevalence of each item in our scale in children of mothers who smoked in pregnancy vs non-smokers, by cohort and gender. All items that have three categories in the scale have been dichotomised. For example, if the prevalence of the 'Restless' behaviours among children of smoker mothers in the BCS cohort is 7.5%, and 5% among non-smoker mothers, the ratio will be 1.5. The error bars represent 95% confidence intervals.

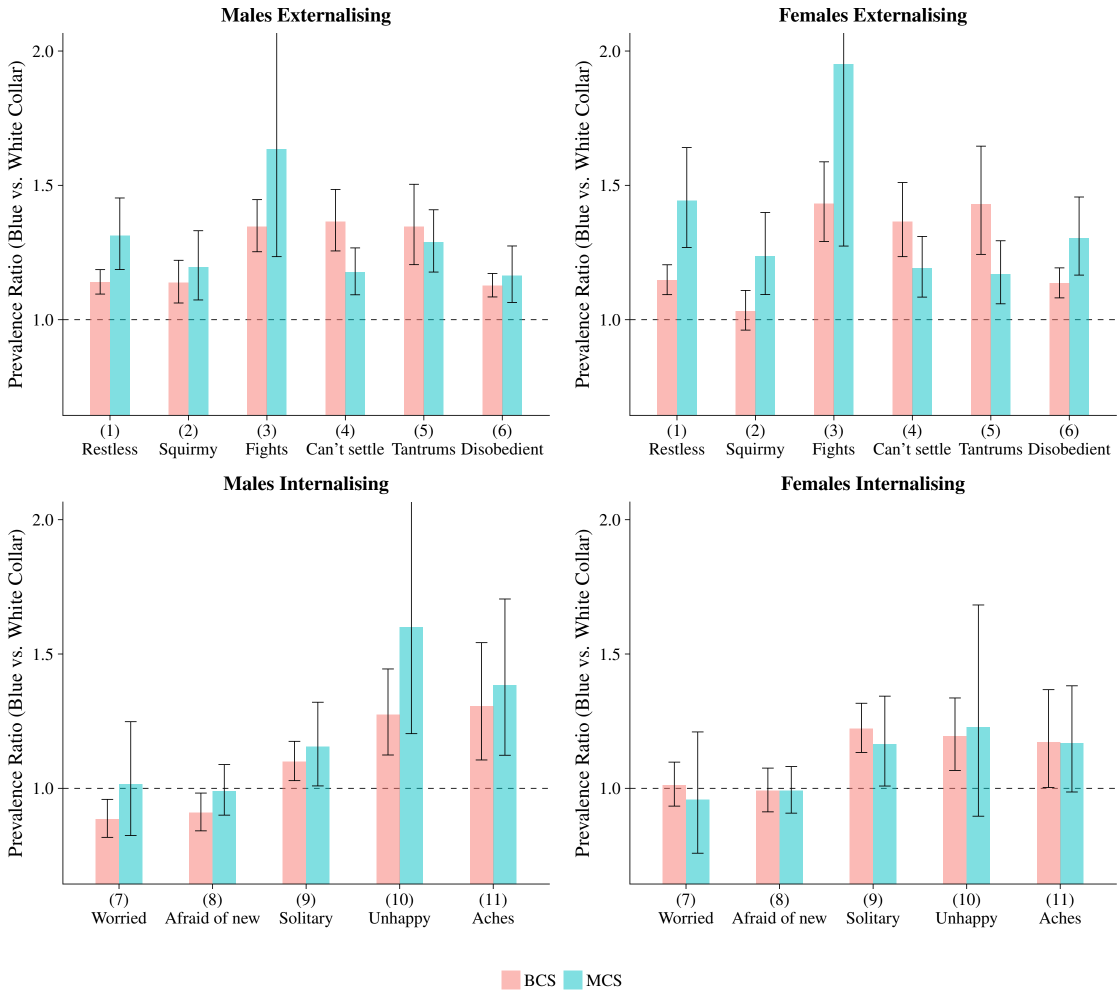

Figure 3.4: Item-level inequality by father's occupation

Notes: The graph displays the ratio between the prevalence of each item in our scale in children of white collar vs blue collar fathers, by cohort and gender. All items that have three categories in the scale have been dichotomised. For example, if the prevalence of the 'Restless' behaviours among children of blue collar fathers in the BCS cohort is 7.5%, and 5% among white collar fathers, the ratio will be 1.5. The error bars represent 95% confidence intervals.

Chapter 4

The role of diet quality and physical activity in the production of adolescent human capital

4.1 Introduction

Adolescent health has risen to prominence in the UK policy debate in the past two decades. Two aspects of well-being have been particularly highlighted: obesity and mental health. Despite levelling off after the mid-2000s, childhood obesity remains at very high levels – with around 20% of 11 year olds being obese in England (NHS, 2018). At the same time, long term secular trends show sharp increases in behavioural and emotional problems similar to obesity (Collishaw et al., 2004, 2010; Collishaw, 2015). Latest estimates show more than one in ten children in England suffering from at least one mental health disorder (NHS, 2017), with recent increases in emotional problems especially for girls (Bor et al., 2014; Fink et al., 2015).

The promotion of diet and physical activity has been suggested as a possible solution to such issues. There are well-known associations between diet and obesity (Jennings et al., 2011; Laverty et al., 2015), and exercise and obesity (Sera et al., 2013; Aars et al., 2018) in early adolescence. Similar associations have been documented between diet and mental health for children in the same age range (see for example the review in O'Neil et al., 2014). Exercise and mental health also increasingly appear to be linked in this development stage (Biddle and Asare, 2011; Griffiths et al., 2016; Ahn et al., 2018).

However, whether a causal relationships actually underlies these associations is not clear (Biddle et al., 2018). This ambiguity is made even starker by the evidence from randomised interventions aiming to foster diet quality and physical activity. Recent metanalyses indicate that the effectiveness of such interventions is at best mixed (Colquitt et al., 2016; Mead et al., 2017). The

This study aims to shed some light on the interplay between obesity, mental health, diet, and exercise. It does so by adopting recent advances in the economics and econometrics of human capital, which allow to explicitly model these relationship within a unified production function framework. A wide consensus has emerged that human capital is a multidimensional concept, encompassing multiple facets from cognition to socio-emotional skills, to physical and mental health (Cunha and Heckman, 2008; Attanasio, 2015). Methodological advances in the estimation of human capital production functions have also enabled more flexible modelling of the complex process of development during early childhood (Cunha et al., 2010; Agostinelli and Wiswall, 2016; Attanasio et al., 2018).

Relatively less attention has been devoted to late childhood and adolescence. However, there is evidence that a second “window of opportunity" exists, with heightened plasticity and sensitivity to the social environment (Blakemore and Mills, 2014; Steinberg, 2015; Fuhrmann et al., 2015). The nature itself of the human capital development process is likely to be different in the adolescent phase. In the preschool years, psychosocial stimulation and nutrition, often decided by parents and carers, seem to play an outsized role (Alderman et al., 2014). In later childhood, the effectiveness of this type of investment fades (Cunha and Heckman, 2008; Del Boca et al., 2014), while the role of school inputs, peers, and the child’s increased agency around time use becomes more prominent (Del Boca et al., 2017).

In this work, I model the production process of human capital in early adolescence, between the ages of 11 and 14. I use the Millennium Cohort Study (MCS), a rich longitudinal survey following the lives of a representative sample of UK children since they were aged 9 months. I focus on two facets of human capital: physical and mental health. The former is measured by body mass, while the latter is measured by parental reports on the Strengths and Difficulties Questionnaire, a commonly used psychometric scale. Two main forms of human capital investment are considered: diet quality and physical activity. In accordance to recent approaches in production function estimation, I do not exclusively rely on directly observed proxy measures for human capital and investments. Rather, I adopt a factor analytic approach to exploit multiple sources of information (where available), explicitly recognising the role of measurement error. To the best of my knowledge, this is the first time that the complementarity between diet and exercise is investigated within a production function framework in the literature. Furthermore, few other studies in the human capital literature have focused on health in adolescence.[1]

Another innovative aspect of this study is the use of novel data sources to deal with the possible endogeneity of human capital investments. Identification exploits the geographical and time variation in the MCS survey. For physical activity, I match respondents to granular weather information on temperature, sunshine, and rainfall. I use time series for local prices of healthy and unhealthy foods, collected for the purposes of computing inflation, to instrument children’s diet. Once interactions and lags are included, this set of price indices becomes potentially high-dimensional. I view this as an optimal instrument problem characterised by approximate sparsity – i.e. where only a relatively small subset of indices might be relevant. Drawing from recent advances in machine learning (Almås et al., 2018), I perform model se-

[1] There is a large literature on the role of time use choices in adolescence – e.g. the effect of time investments by parents or adolescents themselves on subsequent cognition. See Del Boca et al. (2014, 2017) and references therein.

lection for the best subset of instruments via a LASSO procedure (Belloni and Chernozhukov, 2013).

I rely on these sources of exogenous variation to set up a two-step control function strategy that allows me to estimate flexible production functions for human capital. Given that four inputs enter the production function, I test different specifications of a CES, characterised by different combinations of nested inputs. The best-fitting specification is found by nesting previous levels of mental and physical health inside a further CES. Parameter estimates indicate that physical and mental health are complements in production of human capital. Previous levels of human capital are instead highly substitutable with diet quality and physical activity. These results are substantially different from those obtained by using a more restrictive CES specification.

Furthermore, I show that BMI and mental health are very persistent in early adolescence. Endogeneity of diet quality in determining human capital is broadly rejected. A better diet contributes positively to mental health, and does not affect physical health. Physical activity is instead endogenous. After the control function correction, exercise is found to lead to higher BMI. This result is robust to using body fat as a measure of physical health. However, it becomes negative and insignificant in a linear control function specification where a binary indicator for overweight is used as outcome. This indicates that nonlinearities might be at play: exercise increases mean BMI by increasing lean mass in smaller children, but for larger children the effect is null and imprecisely estimated. Significant complementarities emerge between diet and mental health. A higher-quality diet is more effective for children who already have better mental health. Also, improvements in diet from low levels of mental health are more productive.

The rest of the work is structured as follows. Section 4.2 details the human capital production framework used throughout. The main source of data (the MCS survey) is described in Section 4.3. Section 4.4 presents the main challenges to the estimation of the production function, and the empirical strategies used to address them. In particular, Section 4.4.1 sets up the measurement system for the unobserved factors in the analysis, while Section 4.4.2 and Section 4.4.3 elaborate on the instrumental variable strategy and the selection of instruments, respectively. Section 4.5 goes through the main results and contextualises them in the literature. Section 4.6 concludes.

4.2 Production of human capital in adolescence

This work seeks to characterise the human capital production process in early adolescence, namely between the ages of 11 and 14. I focus on two facets of human capital: physical and mental health.[2] I assume that current levels of human capital – physical (θ_t^P) and mental (θ_t^M)

[2] The conceptualisation of mental health I use is very close to what is usually termed *socio-emotional* or *noncognitive skills* in the economics literature (Cunha and Heckman, 2008). In fact, the Strengths and Difficulties Questionnaire is the same instrument that underlies socio-emotional skill in the second chapter of this thesis. In this chapter, given its focus on health production, I prefer the *mental health* denomination.

– are determined by a production process that includes four inputs: previous levels of human capital (θ^P_{t-1}, θ^M_{t-1}), diet quality (θ^D_t), and physical activity (exercise, θ^E_t). In its most general form, the production function can be written as:

$$\theta^k_{i,t} = f_t(\theta^P_{i,t-1}, \theta^M_{i,t-1}, \theta^D_{i,t}, \theta^E_{i,t}, \boldsymbol{X}_{i,t}, A^k_t, \varepsilon^k_{i,t}) \qquad \text{for } k = \{M, P\} \tag{4.2.1}$$

In the above formulation, subscripts $\{M, P\}$ denote physical and mental health, respectively, while $\varepsilon^k_{i,t}$ is an unobserved shock, and A^k_t captures total factor productivity. Finally, $\boldsymbol{X}_{i,t} = \{\boldsymbol{W}_{i,0}, \boldsymbol{W}_{i,t}\}$ is a set of family and child characteristics that might affect the production process. Part of these factors ($\boldsymbol{W}_{i,0}$) are pre-determined with respect to the human capital production process – e.g. measures of family socioeconomic status and child human capital at birth; others ($\boldsymbol{W}_{i,t}$) are contemporaneous to the outcomes, and assumed orthogonal to the unobserved shock term.[3]

Diet quality and physical activity can be considered family investments into the offspring's human capital. At early ages, existing literature assumes that parents are in control of investment choices. Later in childhood and into adolescence, diet and physical activity choices are plausibly determined jointly by the volition of parents and children. In this work, I assume households act as cohesive units in such choices.[4]

4.3 Data

The main source of data for this study is the Millennium Cohort Study (Connelly and Platt, 2014). The MCS is a longitudinal survey following the lives of individuals born across the UK between September 2000 and January 2002, who were alive at 9 months. I take advantage of the longitudinal nature of the study by merging information from multiple waves of the survey: the first wave (when the children were aged around 9 months), the fifth wave (around 11 years of age) and the sixth wave (around 14 years of age). I limit the analysis to non-multiparous children who were living in England or Wales at the sixth wave.[5]

From the first survey, I extract various proxies for family background and socioeconomic status around birth, captured by the $\boldsymbol{W}_{i,0}$ subset of the $\boldsymbol{X}_{i,t}$ vector in Equation 4.2.1. These measures are pre-determined with respect to early adolescence. They include maternal age and education, family income, the child's ethnicity, and whether the household where the child is born comprises a single carer figure. The addition of further SES indicators around birth, such as paternal occupational social class, yields substantially superimposable results. Furthermore, I use birth weight as a proxy for health at birth. This has clear limitations as an

[3] I present separate results for male and female children throughout this study. Gender is included in the $W_{i,0}$ vector whenever males and females are considered jointly.

[4] Observed levels of investments can be rationalised by a unitary utility maximisation model, where the household derives utility from consumption and child human capital. Budget, time, and technical constraints limit the exercise and diet choices that adolescents undertake.

[5] This geographical sample restriction is driven by the availability of weather information – see Section 4.4.2 for details. All data is publicly available at the UK Data Service (University Of London. Institute Of Education. Centre For Longitudinal Studies, 2017a,d,c).

all-encompassing marker of initial health (Conti et al., 2018), but is consistently measured in MCS.

When the cohort members were aged eleven and fourteen, both parents and cohort members were administered a range of survey instruments. I use BMI as the main indicator of a child's physical health. BMI is derived from height and weight measurements taken by enumerators during the household interview. Information on the child's mental health is derived from the main carer's answers to the Strengths and Difficulties Questionnaire (Goodman, 1997, SDQ,). The SDQ is a scale consisting of 25 descriptions of child behaviours that the respondent endorses based on how much they apply to their own child.

The child's diet quality is inferred from a series of questions at the age 14 wave. They are administered directly to the child as part of the interview, and record the frequency of consumption of certain healthy and unhealthy food items such as vegetables, fruit, fizzy drinks, fast food. No direct information on diet quantity is available. Henceforth, references to "diet quality" are related to children's consumption of healthy and unhealthy foods, rather than to the quantity of each of these foods they consume. However, I can't exclude that the diet quality factor is at least partially capturing some aspect of *quantity* as well: some of the diet questions, e.g. the frequency of eating breakfast or consuming fruit and vegetables, might entail a larger calorie intake. As an indicator of physical activity, I use the child's report on the days in the week prior to the interview when they engaged in at least one hour of moderate to vigorous physical activity.[6]

For this work, I limit the analysis to the production of human capital between ages 11 and 14. This is because the exercise and diet information is recorded inconsistently in previous waves, and is less rich than what is available at age 14. Details on all variables from the MCS used in the analysis are available in Table 4.A1. Summary statistics of the sample used in the remainder of the paper are displayed in Table 4.A3.

4.4 Estimation

Estimation of the human capital production function in (4.2.1) presents three main challenges. First of all, human capital and investments are not directly observed in the available data. Secondly, investments might enter endogenously in the production function. Finally, the choice of instrumental variables for investments is subject to a degree of arbitrariness. I tackle each of these issues in what follows.

[6] The 14-year sweep of MCS included a dedicated module measuring physical activity via accelerometer and time use via diaries (Gilbert et al., 2017). While this data is very detailed, its final sample covers less than 5,000 children of the more than 11,000 households in the survey. The relatively small coverage is due to refusals and to respondents not returning the devices. It is thus likely to suffer from significant selection issues – as seen in Rich et al. (2013) for the previous accelerometer sample for the fourth wave of MCS.

4.4.1 Measurement system

In my application, not all the inputs of the production function are perfectly observed. Rather, I observe a set of imperfect measures that contain information about the unobserved constructs of interest (human capital and investments), but are also ridden with measurement error. I thus exploit identification results in Cunha et al. (2010) to infer these latent factors by explicitly specifying a system of measurement equations linking them to what is observed in the data.

Denote as $M^k_{m,j,t}$ the $m = 1, \ldots, N^k$ observed measures for latent factor k for child j in period t. For mental health (at both age 11 and 14), I use $N^M = 20$ items from the Strength and Difficulties Questionnaire (SDQ), administered to the child's mother. The SDQ is designed with reference to five sub-scales pertaining to emotional problems, peer problems, behavioural problems, hyperactivity, and prosocial behaviour. For this work, I exclude the five items belonging to the prosocial scale, and focus on the 20 items that are used to compute the SDQ Total Difficulties score. I assume that the items are measuring a single latent mental health factor. This is a strong simplification, given that the SDQ is formulated to measure multiple facets of mental health (Stone et al., 2010; Goodman et al., 2010). However, a single-factor model still fits the data satisfactorily, and keeps the dimensionality of the production function estimation problem manageable.

For diet quality (at the age 14 wave), I start from a set of eight questions about frequency of consumption of vegetables, fruit, breakfast, sugary drinks, artificially sweetened drinks, fast food, wholemeal bread, and low-fat milk from the child self-completion module. Table 4.A4 shows loadings from an exploratory factor analysis step on the entire set of eight items. The milk fat and sugary drinks indicators do not seem to have a strong correlation with the rest of the diet quality information. I thus drop them from the measurement system, and use the remaining $N^D = 6$ indicators to identify the latent diet quality factor.[7]

For both the mental health and diet quality latent factors, the observed measures are categorical. Each measure M is itself characterised by a continuous latent propensity M^*, which is assumed to have a semi-log relationship with the latent factors:

$$M^{*k}_{m,j,t} = \nu^k_{m,t} + \lambda^k_{m,t} \ln(\theta^k_{j,t}) + \epsilon^k_{m,j,t} \qquad \text{for } k = \{M, D\} \tag{4.4.1}$$

where $\nu^k_{m,t}$ is an item-specific intercept term, $\lambda^k_{m,t}$ are item factor loadings, and $\epsilon^k_{m,j,t}$ are zero-mean measurement errors, and $\{M, D\}$ denote mental health and diet quality respectively. A further set of τ intercept parameters is introduced to discretise this continuous propensity into its categorical observed counterpart:

$$M^k_{m,j,t} = s \qquad \text{if } \tau_{s,m,t} \leq M^{*k}_{m,j,t} < \tau_{s+1,m,t} \qquad \text{for } s = 0, \ldots, N^s_{m,t} \tag{4.4.2}$$

Analogously to an ordered choice model, intercepts ν and thresholds τ cannot be sep-

[7] Some of the items I use to measure diet quality originate from the *eating choices index* (Pot et al., 2014, ECI,). This was developed and validated for adults in the MRC National Survey of Health and Development. It was subsequently included in the age 14 MCS survey. Exploratory factor analysis limited to the items in the ECI reveals low internal consistency – leading to the different choice of indicators for diet used in this work.

arately identified. I thus normalise all intercepts to zero. Furthermore, normalisations are needed to set the scale and location of the latent factors. When continuous measures are available, it is customary to set one measure's loading to unity for each factor, and its intercept to zero, effectively anchoring the factors' unit of measurement – see for example Cunha et al., 2010; Attanasio et al., 2017, 2018). In the categorical case, the measures have no explicit unit of measurement, as they are all defined on a latent dimension. For ease of interpretation, I identify the measurement model by normalising the scale and location of the latent factor, i.e. by setting the mean of the (log) latent variables to zero, and their variance to one. Under standard normality assumptions on latent factors and error terms, the measurement model can be estimated by mean- and variance-adjusted weighted least squares (WLSMV) – see (Muthen et al., 1997).[8]

I define physical health to coincide with body mass, namely log BMI. Height and weight are not self-reported, but objectively measured during fieldwork using scales and meters. It is thus unlikely that BMI suffers from large measurement error. Furthermore, underweight is not prevalent in this sample while overweight and obesity are, making it plausible to consider BMI as having a monotonically decreasing relationship with health.[9] Similarly to physical health, a single, self-reported measure of physical activity frequency is available in the survey data. This prevents from explicitly modelling physical activity as a latent factor. I thus also assume that physical health coincides with the days the child has exercised for more than one hour in the week leading up to the interview.[10]

4.4.2 Endogeneity of investments

Investments in each model period depend on pre-existing child human capital, and other household characteristics – e.g. resources and family composition. The latter are observable dimensions that can be conditioned upon for the analysis. However, endogeneity in investments might arise from unobserved characteristics of the child or of the household. In addition, even in the absence of omitted variables, investments might respond to unobserved shocks to child human capital between model periods. For example, households might react to unexpected episodes of child ill health by adjusting diet and/or physical activity patterns in a compensating or reinforcing manner. Estimates of the production function (4.2.1) that fail to take this into account will be biased.

To address this endogeneity issue, I estimate the production function for human capital using a control function approach. I exploit sources of plausibly exogenous variation in diet quality and physical activity induced by fluctuations in food prices and weather. Under the

[8] All estimates are computed using the `lavaan` package v0.6-2 (Rosseel, 2012).

[9] The use of BMI as a single health indicator has clear limitations (Nevill et al., 2006). However, when age and gender are taken into account, it constitutes a good measure of childhood adiposity (Maynard et al., 2001). Apart from body fat percentage, there are no other objective measures of physical health in the MCS 11 and 14 year surveys. Relying on self-reports from the survey – such as self- or parent-rated health, or longstanding illnesses – would risk conflating physical and mental health. This is avoided by the use of objectively measured height and weight.

[10] Notice that this is equivalent to assuming that the intercept and loading parameters in (4.4.1) are equal to zero and one respectively, that the measurement error ϵ is absent, and that N^P and N^E are both equal to 1.

assumption that households have no control over fluctuations in the prices and weather they face, these are valid instruments for investments.[11] This strategy is possible due to the temporal and geographic variation in the age-14 MCS survey: respondents were interviewed across the entire UK, and in a 15-month period between January 2015 and March 2016. I explicitly write investment functions as follows:

$$\theta^{j}_{i,r,t} = g_t(\theta^{P}_{i,r,t-1}, \theta^{M}_{i,r,t-1}, \boldsymbol{X}_{i,r,t}, \boldsymbol{q}^{j}_{r,t}) \qquad \text{for } j = \{D, E\} \tag{4.4.3}$$

where r indexes individuals by region, and $\boldsymbol{q}^{j}_{r,t}$ are instruments.[12]

To instrument diet quality, I use random variation in prices of healthy and unhealthy food items. I augment the MCS survey dataset with monthly price data from the UK Office of National Statistics (ONS) between 2007 and 2017. This is the same data source as used by ONS to compute the national Consumer and Retail price indices. It constitutes of a large number of nominal price quotes from retailers across the United Kingdom, recorded every month by local price collectors at around 150 locations.

To recover useful exogenous variation from this price dataset, I adopt the following procedure. First, I select price indices for a subset of food items that are most directly related to the MCS survey questions about the child's diet – e.g. price of vegetables, fruit, or takeaway food. Panel A of Table 4.A2 presents the items and the indices used in the analysis.[13] Secondly, I compute the average of each price index by month and Government Office Region. This average is weighted by the item's share in the CPI basket of goods used by ONS. To remove seasonality, I then residualise the time series for each index using month dummies, over the 2007-2017 period. Fourth, I adjust prices for inflation through the food CPI index, and for regional differences in price levels using regional CPI estimates for 2016, available in ONS (2018). Finally, I match these quotes to the month of interview in MCS.[14]

The relative price of healthy and unhealthy foods can influence households' consumption choices.[15] Using the same ONS price data as this study, Jones et al. (2014) show a widening gap between the per-calorie price of healthy and unhealthy foods for the UK in the period 2002-2012. Evidence from scanner data testifies to the sensitivity of household nutrition choices to the price of different foods (Harding and Lovenheim, 2017). Griffith et al. (2015) use rich scanner data for UK households around the 2008 financial crash. They find that changes in food prices resulted in a significant worsening of the nutritional content of households' purchased food. In low-income country settings, many studies have employed heterogeneity

[11] Evidently, households do have control on *where* they live. However, I attempt to isolate the unpredictable component of price and weather variation by subtracting seasonality and adjusting for regional differences (see below).

[12] Notice a slight abuse of notation: the level of geographical granularity (r) is different between prices (Government Office Region) and weather (Output Area).

[13] Alongside items related to the survey questions, I also include indices for items that do not have a direct counterpart in the survey, such as crisps and sweets.

[14] Given that the day of interview is not accessible in the MCS dataset, this match is necessarily somewhat rough. The price quote data can be freely obtained from the ONS website at `https://www.ons.gov.uk/economy/inflationandpriceindices`.

[15] See (Cawley, 2015) for a brief review on the link between food prices and obesity.

in local food prices to identify the effect of diet and nutrition on child development – see for example Puentes et al. 2016; Attanasio et al. 2017, 2018.

For physical activity, I use random variation in weather patterns.[16] I employ a novel dataset containing weather data for the whole of the UK in the years between January 2006 and December 2016 (Hollis and McCarty, 2017). This data consists of monthly observations of four climatic variables – temperature, rainy days, rainfall, and sunshine time – at a resolution of 5 km by 5 km grid points (Perry and Hollis, 2005).

To generate the instrument, I first overlay the grid points onto a map of granular statistical subdivisions (Output Areas, OAs), used for the UK census and for ONS neighbourhood statistics. Each OA is matched to the nearest weather grid point using its centroid. Secondly, I residualise the time series for each weather variable through month dummies across the 11 year period considered. The residualisation is carried out *within* each OA. The aim of this step is is to remove seasonality, so that the instrument only retains fluctuations away from the long-term average weather in each small geographical unit. Finally match MCS respondents to Output Areas using their location as recorded in the 14 year survey.[17]

The relationship between weather and physical activity in children is well documented in a variety of contexts (Tucker and Gilliland, 2007; Duncan et al., 2008; Bélanger et al., 2009; Harrison et al., 2017). Findings from the SPEEDY study in Norfolk, UK confirm a robust association between rainfall and exercise measured with accelerometers (Harrison et al., 2011, 2015). Similarly to food prices, weather has been used as an instrumental variable for other relationships. Jacob et al. (2007) use anomalous temperature and precipitation to explore autocorrelation in crime rates. In their US application on PSID data, Laidley and Conley (2018) employ geographical and seasonal variation in sunlight to disentangle the causal effect of active and passive leisure time on cognition.

4.4.3 Instrument selection

The set of nine price indices used to instrument diet quality is informed by the survey questions underlying the latent diet quality factor. However, the way in which they enter the investment function is a priori unknown. There might be lagged effects, where variation in prices for previous months affect current consumption, or interaction effects between different indices. The goal of these instrumental variables, conditional on their validity, is to explain the most variation in the endogenous inputs to the production function. This can be viewed as an *optimal instrument* problem, where a large number of plausible instruments is available and the identity of the relevant instruments is a priori unknown.

To select relevant price instruments amongst all contemporaneous prices, their lags, and the corresponding two-way interactions, I follow a least absolute shrinkage and selection op-

[16]The use of weather data as a source of exogenous variation is increasingly popular in economics (Dell et al., 2014).

[17]At the time of writing, only data for England and Wales had been processed for use, which is the reason why the analysis is limited to these two countries. More details and variable definitions are available in Panel B of Table 4.A2.

erator (LASSO) approach, as put forward by Belloni and Chernozhukov (2013). Denote $\boldsymbol{Q}^{j}_{r,t}$ as the p-dimensional set of possible instruments for diet quality. This includes prices in the same month of the interview, up to 3 monthly lags, and all interactions. I start from a log-linear approximation to the investment function g_t in (4.4.3):[18]

$$\log \theta^{D}_{i,r,t} = \mu_1 \log \theta^{P}_{i,r,t-1} + \mu_2 \log \theta^{M}_{i,r,t-1} + \boldsymbol{X}_{i,r,t}\boldsymbol{\mu}_3 + \boldsymbol{Q}^{D}_{r,t}\boldsymbol{\beta} + v^{D}_{i,r,t} \tag{4.4.4}$$

As a first step, I partial out lagged human capital and covariates from the left-hand side, and express the above relationship as:

$$\log \tilde{\theta}^{D}_{i,r,t} = \boldsymbol{Q}^{D}_{r,t}\boldsymbol{\beta} + \tilde{v}^{D}_{i,r,t}$$

The Lasso estimator for $\boldsymbol{\beta}$ is based on minimising the same squared residuals objective function as OLS, but it adds a penalty term for each additional regressor:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \left[\left(\log \tilde{\theta}^{D}_{i,r,t} - \boldsymbol{Q}^{D}_{r,t}\boldsymbol{\beta} \right)^2 \right] + \frac{\lambda}{n} \parallel \hat{\boldsymbol{\Psi}}\boldsymbol{\beta} \parallel$$

where $\parallel \cdot \parallel$ denotes the l_1 norm, and $\boldsymbol{\Psi} = \text{diag}(\hat{\psi}_1, \ldots, \hat{\psi}_p)$ is the penalty matrix. The procedure identifies a lower-dimensional set of instrumental variables $\boldsymbol{q}^{D}_{r,t}$ with non-zero coefficients in (4.4.4). This subset contains the variables with the highest predictive power among the larger set of instruments $\boldsymbol{Q}^{D}_{r,t}$.[19] The smaller set $\boldsymbol{q}^{D}_{r,t}$ is finally the one used in the OLS first stage of the control function estimation of the production function. This is in fact equivalent to a Post-Lasso step, albeit with no selection on the other covariates.

Only four weather indicators are available to instrument physical activity (Panel B of Table 4.A2). The optimal instrument approach used for food prices is known to perform less well if the relationship is not approximately sparse – i.e. if there are relatively many relevant instruments in the set that is considered. I thus manually select weather instruments for physical activity. Rainfall in the month of interview (measured in millimetres) exhibits the strongest association with the reported frequency of physical activity in children, and will serve as the instrument for physical activity in the remainder of this study. Consistently with the idea that contemporaneous weather conditions are what influences exercise, lags and leads of weather measures do not impact physical activity recorded in a given month.[20]

[18]A linear approximation (without logs) yields substantially equivalent results, but explains a lower fraction of the variation in the investments. Results are available upon request.

[19]The *rigorous* Lasso estimator is implemented in the R package `hdm` (Chernozhukov et al., 2016). The penalisation parameters λ and $\boldsymbol{\Psi}$ are not chosen by cross-validation, but instead set to theoretically justified values – hence the *rigorous* appellative. These values are $\hat{\psi}_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Q}^{D}_{j,r,t}\hat{\tilde{v}}^{D}_{i,r,t}}$ and $\lambda = 2c\sqrt{n}\boldsymbol{\Phi}^{-1}(1-\gamma/(2p))$, where the constant c is set to .5 and γ is a confidence level set to .1. Heteroskedasticity is incorporated by the $\hat{\tilde{v}}^{D}_{i,r,t}$ term in the penalties, using an iterative procedure (Belloni et al., 2012).

[20]Additional results are available upon request.

4.4.4 Specification of the production function

Even though nonparametric identification is possible, estimation of the production function (4.2.1) requires a parametric approximation. In line with previous research, this work adopts a constant elasticity of substitution (CES) specification. The CES has the advantage of allowing for nonlinear complementarities between inputs in a parsimonious way, while at the same time delivering interpretable parameter estimates. A standard CES for the problem at hand can be written as:

$$\theta_{i,t}^{k} = A_t^k \left[\gamma_1^k \left(\theta_{i,t-1}^{P}\right)^{\rho_k} + \gamma_2^k \left(\theta_{i,t-1}^{M}\right)^{\rho_k} + \gamma_3^k \left(\theta_{i,t}^{D}\right)^{\rho_k} + \gamma_4^k \left(\theta_{i,t}^{E}\right)^{\rho_k}\right]^{\frac{\phi_k}{\rho_k}} \cdot \exp\left\{\boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varepsilon_{i,t}^{k}\right\} \ , \quad \sum_{l=1}^{4} \gamma_l^k = 1. \tag{4.4.5}$$

Non-constant (increasing or decreasing) returns to scale are captured by the ϕ_k parameter, and complementarity between inputs by ρ_k. It nests a linear production function with perfectly separable inputs for $\rho_k = 1$, and the Cobb-Douglas case when $\rho_k = 0$. However, it is quite restrictive, as it constrains the four inputs to have the same degree of complementarity with one another.

More flexible patterns of substitutability between inputs can be achieved by nesting additional CES functions within the main function (Attanasio et al., 2017). This frees further combinations of inputs to exhibit differential complementarity. A natural choice would be to additionally nest diet quality and physical activity, or lagged physical and mental health. An even richer specification would “double nest" lagged human capital and investments separately. In this work, we test the following possible specifications of the CES:

- Nested investments CES

$$\theta_{i,t}^{k} = A_t^k \left[\gamma_1^k \left(\theta_{i,t-1}^{P}\right)^{\rho_k} + \gamma_2^k \left(\theta_{i,t-1}^{M}\right)^{\rho_k} + \gamma_3^k \left(\delta_1^k \left(\theta_{i,t}^{D}\right)^{\rho_k^I} + \delta_2^k \left(\theta_{i,t}^{E}\right)^{\rho_k^I}\right)^{\frac{\rho_k}{\rho_k^I}}\right]^{\frac{\phi_k}{\rho_k}} \cdot \exp\left\{\boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varepsilon_{i,t}^{k}\right\} \tag{4.4.6}$$

  with $\gamma_1^k + \gamma_2^k + \gamma_3^k = 1$, $\delta_1^k + \delta_2^k = 1$.

- Nested lags CES

$$\theta_{i,t}^{k} = A_t^k \left[\gamma_1^k \left(\delta_1^k \left(\theta_{i,t-1}^{P}\right)^{\rho_k^L} + \delta_2^k \left(\theta_{i,t-1}^{M}\right)^{\rho_k^L}\right)^{\frac{\rho_k}{\rho_k^L}} + \gamma_2^k \left(\theta_{i,t}^{D}\right)^{\rho_k} + \gamma_3^k \left(\theta_{i,t}^{E}\right)^{\rho_k}\right]^{\frac{\phi_k}{\rho_k}} \cdot \exp\left\{\boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varepsilon_{i,t}^{k}\right\} \tag{4.4.7}$$

  with $\gamma_1^k + \gamma_2^k + \gamma_3^k = 1$, $\delta_1^k + \delta_2^k = 1$.

- Double nested CES

$$\theta_{i,t}^{k} = A_t^k \left[\gamma_1^k \left(\delta_1^k \left(\theta_{i,t-1}^{P}\right)^{\rho_k^L} + \delta_2^k \left(\theta_{i,t-1}^{M}\right)^{\rho_k^L}\right)^{\frac{\rho_k}{\rho_k^L}} + \gamma_2^k \left(\delta_3^k \left(\theta_{i,t}^{D}\right)^{\rho_k^I} + \delta_4^k \left(\theta_{i,t}^{E}\right)^{\rho_k^I}\right)^{\frac{\rho_k}{\rho_k^I}}\right]^{\frac{\phi_k}{\rho_k}} \cdot \exp\left\{\boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varepsilon_{i,t}^{k}\right\} \tag{4.4.8}$$

  with $\gamma_1^k + \gamma_2^k = 1$, $\delta_1^k + \delta_2^k = 1$, $\delta_3^k + \delta_4^k = 1$.

Notice that the nested CES functional forms provide straightforward specification tests through the global and nested complementarity parameters. For example, in (4.4.6), if $H_0 : \rho_k - \rho_k^I = 0$ is not rejected, then the nesting does not improve the function's fit to the data. This is because the ρ_k/ρ_k^I term becomes unity, and the nested CES becomes equivalent to its non-nested alternative. Similarly, $H_0 : \phi_k = 1$ is the testable null hypothesis of constant returns to scale. However, increased flexibility comes at a cost. The nonlinear least squares estimator fails to reach convergence more often when more complex functional forms are estimated. Furthermore, allowing for non-constant returns to scale ($\phi_k \neq 1$) tends to render the estimation of the complementarity parameter ρ_k less stable.

4.4.5 Estimation steps

To sum up, estimation of the production function in (4.2.1) is carried out in the following steps:

1. The measurement system in (4.4.1) is estimated, and factor scores for each child are computed based on the estimated measurement parameters.
2. The optimal set of instruments for diet quality is identified using Lasso.
3. The human capital production function is estimated by nonlinear least squares, using a control function approach: residuals from the two first stage models for diet quality and physical activity are plugged into the main equation alongside covariates $\boldsymbol{X}$. Inference for this step is bootstrap-based.[21]

4.5 Results

This section is organised as follows. I first briefly present the estimates from the measurement system and the investment functions. Secondly, I test different CES functional forms for the human capital production function. Finally, I present the estimates of the production function and analyse the implications for the marginal products of diet quality and exercise.

Measurement system

Table 4.A5 displays estimated loadings and informational content for the measurement system in (4.4.1). The SDQ items measuring mental health (both at 11 and and 14 years of age) load satisfactorily on the latent factor, and exhibit good signal-to-noise ratios. They are administered at the same time during the parental interview, and they all stem from a separately validated psychometric instrument – the Strengths and Difficulties Questionnaire – which might explain their high internal consistency and informational content. The same is true for the diet quality indicators. While the frequency of breakfast and the type of bread consumed exhibit slightly smaller loadings, they are mostly above the commonly accepted threshold of .4. No particular differences between male and female children emerge.

[21] The main results are obtained by simply inserting the estimated residuals linearly in the total factor productivity. Estimates obtained using more complex functions of these residuals, like second- or third-order polynomials, are almost identical and are available from the author on request.

Investment functions

Next I examine the estimates for the log-linear approximations to the investment equations in (4.4.4), reported in Table 4.1. In my specification, diet quality and physical activity at age 14 depend on lagged human capital, socioeconomic status at birth, and exogenous shifters – food prices for diet, and weather for exercise. Body mass at the previous interview is negatively associated with physical activity, but not diet. Actually, there is a positive relationship between diet quality and previous BMI. As highlighted in Section 4.3, the diet quality factor might be capturing quantity as well – via the breakfast, fruit, and veg frequency indicator. Children with better mental health have both a significantly better diet and exercise more. Socioeconomics status at birth (both measured by maternal education and income) predicts diet, but not exercise. Birthweight, a measure for health at birth, is not significantly associated with either investment. This can be explained by birthweight being highly correlated with subsequent childhood BMI.

In general, as evidenced by the adjusted R^2 values, the first stages have fairly low explanatory power for physical activity. This is because there is just a single measure of physical activity for the age 14 sample, and thus I cannot exploit multiple sources of information to deal with measurement error. The result underscores the importance of measurement issues in the estimation of human capital production functions (Cunha et al., 2010). Relatedly, the F-statistics for the food price instruments are comfortably above the conventional relevance thresholds (Stock and Yogo, 2005), while the weather instrument appears weaker, especially for the sample of females.

CES specification tests

As highlighted in Section 4.4.4, both nesting and returns to scale are testable from linear combinations of estimated parameters. The bootstrap inference used for confidence intervals lends itself straightforwardly to the computation of CIs for the specification tests as a byproduct of the inference from the main CES parameters. Table 4.2 presents estimates and confidence intervals for such specification tests.

Panel A presents tests for the nesting level, against the plain CES alternative where all inputs have the same degree of substitutability. Recall from Section 4.4.4 that rejecting that the difference between the global and nested complementarity parameter is zero indicates that the additional nesting is not providing a better fit to the data. The specification with nested investments produces wide confidence intervals, and has low convergence rates (see the bottom line of Table 4.A7). Nesting lagged levels of human capital instead results in estimates of $\rho - \rho^L$ that are away from zero, particularly in the production of physical health. In the case of mental health, estimates of the difference are more noisy. Nesting both investments and lags leads to similar conclusions: allowing previous levels of human capital to exhibit different complementarity than the rest of the inputs improves the fit. I thus rely on the *nested lags* CES in (4.4.7) as my preferred specification.

Returns to scale (RTS) are tested in Panel B of Table 4.2. Estimates of the ϕ parameter for the plain CES specification would indicate decreasing RTS for BMI and increasing RTS for mental health. However, estimates of ϕ in the preferred nested lags specification are much closer to (and often indistinguishable from) unity. Moreover, freeing the additional RTS parameter renders estimation of the ρ complementarity parameter less precise. For this reason, I will focus on results keeping $\phi_k = 1$.

Production function estimates

Parameter estimates for the nested lags CES specification from (4.4.7) are presented in Table 4.3. Here, lagged physical and mental health are nested into a single composite. This “lagged health composite" enters the outer CES together with diet quality and physical activity. The units of measurements for the inputs are as follows: physical health is measured in log units of BMI, while mental health, diet quality, and physical activity are expressed in standard deviations.[22]

The coefficients on the control functions (the residuals from the first stage, $\hat{v}^D$ and $\hat{v}^E$) are denoted by φ^D and φ^E. They serve as a straightforward test of the endogeneity of investments, in the spirit of a variable addition test (Wooldridge, 2014). Diet quality does not seem to be endogenous in either the production function of physical or mental health. The coefficient on the physical activity residual in the BMI equation is instead significant and negative across gender groups, indicating the presence of endogeneity.

Consider first the production function for physical health, in the leftmost panel of Table 4.3. BMI is estimated to be highly persistent. However, there is significant cross-productivity of mental health: better mental health at age 11 results in lower BMI at 14. Diet quality has a small and barely significant effect on males' BMI. It is possible this is because the diet quality factor is also capturing calorie intake to some degree. However, exercise has a large and positive effect.

This positive effect of exercise on BMI is not an artifice of CES estimation. Table 4.A9 presents estimates from both CES and linear production functions, for the whole sample of children, controlling for gender. It compares control function specifications, that seek to correct for potential endogeneity of investments, with exogenous specifications, where no endogeneity correction is applied. In the exogenous version, there is barely any association between physical activity and BMI. Once the control function is included, the coefficient on physical activity becomes positive and significant.[23]

It is possible that this average effect on BMI is masking significant heterogeneity. In that case, children with lower body mass might be driving the estimate, for whom more frequent

[22] I will focus on the nested lags CES specifications. As highlighted in Section 4.4.4, previous levels of human capital are nested into a “composite". The parameter γ_1 captures global productivity of the lagged human capital composite, while δ_1 and δ_2 allow the productivity of each component of human capital to be different from the overall coefficient on the composite. Estimates for other functional forms are available for completeness in Tables 4.A6, 4.A7, and 4.A8.

[23] The sign of the effect of physical activity is also not confined to BMI as a measure of body composition. Panel A of Table 4.A10 replicates the linear production function in Table 4.A9, but uses log body fat percentage instead of BMI as a measure of physical health. Just like with BMI, the negative sign reverses when accounting for endogeneity of exercise – although it is not statistically significant.

exercise would increase lean mass rather than decrease fat mass. To shed more insight on this, Panel B of Table 4.A10 investigates possible nonlinearities of the relationship between body size and physical activity, by expressing physical health with a dummy for overweight. In this case, the sign on physical activity does not reverse – even if it becomes insignificant. It is not possible to replicate this result allowing for CES-type complementarities, since the CES is not defined for discrete inputs.

It is useful to frame this result in the context of the biomedical and epidemiological literature on the association between physical activity and body mass. This relationship has been documented in a variety of contexts (Janssen et al., 2005). However, the association tends to be stronger when looking at body composition rather (fat/lean mass) rather than BMI (Reichert et al., 2015; Ramires et al., 2016; Aars et al., 2018). This is true in the MCS data used here, as evidenced by a significant negative coefficient of PA on body fat percentage in the OLS estimates in Table 4.A10, while the association with BMI in Table 4.A9 is barely significant.

The way in which physical activity is measured plays an important role in the detection of effects. Associations are not always apparent when using self-reported measures like in this work (Wareham et al., 2005), but tend to be stronger when objectively measured data is used. For example, strong negative gradients between accelerometer-derived measures of PA and body mass are observed in Ness et al. (2007) for the ALSPAC study, and in Sera et al. (2013) for the MCS study, among others. Accelerometer data is indeed available for MCS adolescents in the age 14 wave. However, there are at least three main reasons that caution against relying on such measures. First, the low response rates are likely to produce a very selected sample – as documented by Rich et al. (2013) for the age 7 accelerometer study in MCS. Secondly, the reduced sample size (around half of what is available here) renders estimation of complex production functions more difficult. Finally, a smaller sample size would also render my instrumental variable strategy less viable, reducing the relevance of the instruments.

With some exceptions (Jennings et al., 2011; Hamer and Stamatakis, 2018), epidemiological studies do not model the dynamic persistence of body mass or composition, being limited to cross-sectional associations. Even less common are attempts to go beyond correlations and estimate causal effects. Some causal evidence is available from evaluation of interventions. A recent Cochrane review on the effect of family-based behaviour changing RCTs (targeting diet and/or physical activity) in children finds small, short-term reductions in BMI. The evidence from these interventions is deemed to be of low quality (Mead et al., 2017). The effectiveness of school interventions promoting physical activity is also in doubt: the meta-analysis in Harris et al. (2009) finds no effect on BMI.

Parameters related to the production of mental health are displayed in the rightmost panel of Table 4.3. Just like BMI, mental health is markedly persistent between ages 11 and 14 – or self-productive, in the terminology of Cunha and Heckman (2007). However, there seems to be no cross-productivity of BMI on mental health. A better quality diet has significant effects on mental health, especially for females, while the effect of physical activity is small and imprecisely estimated across all groups.

This result is in line with the existing literature (O'Neil et al., 2014). Using a behaviour-based measure of mental health similar to this study, Oddy et al. (2009) show a positive gradient between fruit and vegetable consumption and scores on the Child Behavioural Checklist for a sample of Australian adolescents. Evidence on the relationship between mental health (measured by the SDQ) and junk food or sweets is available for the UK (Wiles et al., 2009) and Germany (Kohlboeck et al., 2012).

Complementarity between inputs

The CES specification allows for nonlinear patterns of complementarity between inputs, captured in a parsimonious way by the complementarity parameter ρ. The nested lags specification in (4.4.7) adds an additional CES layer, with a further complementarity parameter ρ^L. Table 4.3 presents estimates of the complementarity parameters and of the corresponding elasticities of substitution – i.e. $1/(1-\rho)$.

For physical health, there is more pronounced complementarity in the nested CES, indicating that BMI and mental health in the previous period are complements in the production of next-period BMI. The implied degree of complementarity is negative, but approximately equivalent to Cobb-Douglas ($\rho^L = 0$). The outer CES, where the inputs are a composite of lagged human capital, diet quality, and exercise, is characterised by almost perfect substitutability, with ρ approaching unity. For mental health, estimates of the complementarity parameters are noisier. Still, it holds true that the inner CES seems to exhibit a higher degree of complementarity.

It is helpful to visualise the marginal products corresponding to the estimated CES parameters. In Figure 4.1 and Figure 4.2, I present the marginal products (MPs) of diet quality and physical activity on physical (Panel A, top) and mental health (Panel B, bottom), separately for males and females. Each sub-panel shows MPs alongside the distribution of the four inputs to the production function, while the other inputs are held constant at their mean level. Here, the slope of the lines conveys the degree of complementarity.

As seen in Table 4.3, diet, exercise, and human capital are highly substitutable in the production function for BMI. This is testified by the relative flatness of the marginal products, for both males and females. However, more relevant complementarity emerges for mental health. Diet quality appears to be more productive at higher levels of mental health, indicating significant complementarity. The MP of diet on mental health is also decreasing along its own distribution, suggesting that there are higher gains to diet improvements for children with worse initial diet. A possible policy implication is that policies promoting a healthier diet in disadvantaged children will have comparatively high returns on mental health, while not affecting BMI. As expected, estimated complementarities are less pronounced when the CES specification is less flexible.

4.6 Conclusion

This work provides an innovative application of production function estimation to a topical question, namely the role played by physical activity and diet in health outcomes during adolescence. I investigate the dynamics of human capital accumulation between the age of 11 and 14, in a sample of children residing in England and Wales. Despite evidence that this period is relevant for later life outcomes, it is relatively understudied. Furthermore, I shift the human capital analysis away from its traditional focus on cognition towards the production of health. Multiple measures of mental health and diet quality are combined in a factor analytic framework, which takes into account measurement error. Exploiting novel data sources providing exogenous variation in diet and physical activity, I'm able to disentangle their effects in the production of physical and mental health using a control function approach.

A first important finding, as has been recently highlighted in Attanasio et al. (2017), is the importance of allowing for flexibility in the specification of the production function. In my application, a simple CES unduly restricts the possible complementarity patterns between the four inputs. The best fit is achieved by nesting an additional CES which allows previous levels of human capital to be more complementary between each other than the rest of the inputs.

While diet quality does not seem to enter the production process endogenously, I find physical activity to be endogenous in the production process of physical health. After endogeneity correction by control functions, my findings reveal a positive effect of physical activity on BMI. The sign of the effect is robust to the adoption of a simpler linear specification, and is similar to what is obtained using body fat rather than BMI. This average effect is likely masking considerable heterogeneity, with smaller children seeing physical activity increase their body mass. In general, the production of BMI does not seem to exhibit significant levels of complementarity between starting levels of human capital and the diet and exercise investments considered.

For mental health, diet quality seem to play a consistent positive role – in accordance with existing observational literature. The positive effect is more pronounced for children with a poorer diet. This is potentially relevant for policy, given that a better diet might lead to improvements in mental health for children at the worst end of the diet distribution, who are plausibly poorer.

Some limitations of this study must be underlined. It uses a single survey measure of physical activity concerning frequency of moderate to vigorous exercise, which is also self-reported. This relatively coarse and error-ridden measure renders the analysis more imprecise, and might mask other margins not capture by frequency alone. Similarly, the use of BMI as a single indicator of physical health is not ideal. Even if measurement error is a less relevant issue in this objectively reported measure, body mass might exhibit heterogeneous and nonlinear relationships with diet and exercise which cannot be captured in a CES framework. Finally, the specification I adopt, where the outcome of the production process and the investments are contemporaneous, is more prone to issues of reverse causation.

4.7 Tables

Table 4.1: Investment function estimates

	Diet Quality (14)			Physical Activity (14)		
	All	Males	Females	All	Males	Females
BMI (11) – $\log\theta^P_{t-1}$	0.037	0.176*	−0.118	−0.187***	−0.180***	−0.193***
	(0.064)	(0.090)	(0.090)	(0.033)	(0.048)	(0.047)
Mental health (11) – $\log\theta^M_{t-1}$	0.228***	0.202***	0.252***	0.044***	0.048***	0.039***
	(0.013)	(0.018)	(0.019)	(0.007)	(0.010)	(0.010)
Rural resident (11)	0.132***	0.108***	0.144***	0.009	−0.011	0.029
	(0.024)	(0.034)	(0.034)	(0.012)	(0.018)	(0.018)
Mother ed. – NVQ 2 (GCSEs)	0.120***	0.116**	0.124***	0.009	0.025	−0.008
	(0.034)	(0.048)	(0.047)	(0.019)	(0.027)	(0.027)
Mother ed. – NVQ 3 (A Level)	0.236***	0.255***	0.204***	0.022	0.028	0.015
	(0.040)	(0.056)	(0.056)	(0.022)	(0.031)	(0.031)
Mother ed. – NVQ 2 (Higher Ed.)	0.445***	0.397***	0.492***	0.045**	0.054**	0.033
	(0.034)	(0.049)	(0.048)	(0.019)	(0.027)	(0.027)
Mother ed. – Other qualif.	0.290***	0.374***	0.196**	0.095***	0.120**	0.081
	(0.073)	(0.107)	(0.097)	(0.037)	(0.051)	(0.053)
Family inc. $\leq$ 60% median	−0.127***	−0.151***	−0.113***	−0.004	0.025	−0.033
	(0.030)	(0.043)	(0.042)	(0.016)	(0.023)	(0.023)
Birthweight (kg)	0.014	0.043	−0.019	0.005	0.001	0.010
	(0.019)	(0.027)	(0.028)	(0.010)	(0.014)	(0.014)
Female	0.041*			−0.133***		
	(0.021)			(0.011)		
Food price indices						
Eggs * Sweets	0.001***					
	(0.000)					
Takeaway * Sweets (L3)	−0.001***					
	(0.000)					
Breakfast cereal	0.019**					
	(0.008)					
Breakfast cereal * Sweets	0.001**					
	(0.001)					
Takeaway (L3) * Crisps (L2)			0.003***			
			(0.001)			
Takeaway (L3) * Sweets (L3)			−0.001			
			(0.001)			
White bread * Breakfast cereal		−0.001				
		(0.000)				
White bread * Eggs (L1)		0.001***				
		(0.000)				
Weather						
Rainfall (mm) in month of int. (de-seasonalised)				−0.001***	−0.001***	−0.001**
				(0.000)	(0.000)	(0.000)
Adj. R^2	0.209	0.183	0.239	0.053	0.023	0.033
Joint F-stat of instruments	17.606	17.553	25.854	13.244	10.036	4.284
Observations	4802	2447	2355	4802	2447	2355

Notes: The table reports OLS coefficients and heteroskedasticity-robust standard errors (in parentheses) from log-linear approximations to the investment equations in (4.4.4). Significance: * (10%), ** (5%), *** (1%) Additional controls (not shown) are for mother age at birth, non-white ethnicity, children born to single parents, and number of siblings and season of interview at age 14 (January-March is winter, April-June is spring, and so on). Price indices are cleaned of seasonality, adjusted for food inflation and regional price differences, and normalised to mean 100. (LN) denotes lagged prices of N months.

Table 4.2: Specification tests

CES Functional form		BMI All	BMI Males	BMI Females	Mental health All	Mental health Males	Mental health Females
Panel A: Nesting tests							
Nested invest.	$\rho - \rho^I$	0.209 [−1.100, 1.919]	−0.461 [−2.171, 1.518]	0.746 [−0.870, 4.360]	0.484 [−0.764, 32.696]	0.342 [−23.438, 15.463]	1.486 [−0.490, 27.401]
Nested lags	$\rho - \rho^L$	**0.932** [0.709, 1.159]	**0.664** [0.364, 0.929]	**1.204** [0.854, 1.632]	−0.483 [−0.828, 4.797]	1.730 [−0.527, 11.497]	3.053 [−0.821, 4.377]
Double nested	$\rho - \rho^L$	**0.943** [0.703, 1.170]	**0.635** [0.323, 0.938]	1.179 [−17.182, 1.565]	**3.880** [2.226, 5.716]	1.993 [−0.453, 10.185]	**3.169** [2.012, 4.307]
	$\rho - \rho^I$	0.561 [−0.714, 1.683]	−0.174 [−1.326, 0.777]	0.105 [−1.174, 3.540]	0.617 [−0.872, 10.268]	0.057 [−26.341, 9.871]	0.745 [−0.547, 9.682]
Panel B: Returns to scale tests							
Plain	ϕ	**0.865** [0.839, 0.900]	**0.919** [0.851, 0.966]	**0.834** [0.790, 0.927]	**1.240** [1.146, 1.383]	**1.131** [1.030, 1.371]	**1.355** [1.200, 1.602]
Nested invest.	ϕ	**0.864** [0.839, 0.902]	**0.921** [0.848, 0.967]	**0.854** [0.795, 0.935]	**1.259** [1.134, 1.452]	**1.151** [1.037, 1.415]	**1.366** [1.184, 1.604]
Nested lags	ϕ	**0.906** [0.840, 0.948]	**0.949** [0.845, 1.006]	**0.895** [0.811, 0.949]	**0.965** [0.714, 1.422]	**0.988** [0.525, 1.530]	**0.878** [0.565, 1.426]
Double nested	ϕ	**0.909** [0.861, 0.947]	**0.949** [0.865, 1.006]	**0.891** [0.809, 0.961]	**1.301** [1.155, 1.509]	**1.259** [1.038, 1.573]	**1.375** [1.193, 1.653]

Notes: The table presents linear combinations of estimates of the parameters of CES production functions. Panel A shows estimated differences in complementarity coefficients from (4.4.6), (4.4.7), and (4.4.8), with the returns to scale parameter ϕ_k fixed to unity. Panel B shows estimates for the returns to scale parameter ϕ_k from (4.4.6), (4.4.7), and (4.4.8). They are obtained using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged in to the second stage CES equation. The CES function is estimated by nonlinear least squares using the Levenberg-Marquardt algorithm (Elzhov et al., 2016). The control function procedure is bootstrapped with 1000 repetitions: medians from the bootstrapped distribution are on top, and 95% confidence intervals are in brackets. *N*: 4,802 (all), 2,447 (males), 2,355 (females). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

Table 4.3: CES Health production functions – Nested lags

$$\theta^k_{i,t} = A^k_t \left[\gamma^k_1 \left(\delta^k_1 \theta^{P}_{i,t-1}{}^{\rho^L_k} + \delta^k_2 \theta^M_{i,t-1}{}^{\rho^L_k}\right)^{\rho_k/\rho^L_k} + \gamma^k_2 \left(\theta^D_{i,t}\right)^{\rho_k} + \gamma^k_3\left(\theta^E_{i,t}\right)^{\rho_k}\right]^{\frac{1}{\rho_k}} \cdot \exp\left\{\boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varphi^D_k \hat{v}^D_{i,t} + \varphi^E_k \hat{v}^E_{i,t} + \varepsilon^k_{i,t}\right\} \quad \text{with} \quad \gamma^k_1+\gamma^k_2+\gamma^k_3 = 1, \delta^k_1+\delta^k_2=1.$$

	log BMI – θ^P_t			**Mental health** – θ^M_t		
	All	Males	Females	All	Males	Females
Lagged hum. capital composite (γ_1)	**0.794** [0.747, 0.835]	**0.847** [0.789, 0.903]	**0.712** [0.642, 0.768]	**1.051** [0.846, 1.171]	**0.904** [0.859, 1.252]	**0.825** [0.788, 1.136]
Diet quality (γ_2)	0.003 [−0.003, 0.009]	**0.008** [0.000, 0.017]	−0.003 [−0.010, 0.005]	**0.085** [0.012, 0.149]	0.072 [−0.015, 0.145]	**0.134** [0.043, 0.202]
Physical activity (γ_3)	**0.202** [0.159, 0.252]	**0.144** [0.091, 0.202]	**0.291** [0.232, 0.363]	−0.115 [−0.245, 0.050]	0.010 [−0.284, 0.074]	0.035 [−0.228, 0.096]
log BMI (δ_1)	**1.007** [1.004, 1.010]	**1.008** [1.005, 1.012]	**1.006** [1.003, 1.011]	0.069 [−0.048, 0.157]	−0.017 [−0.064, 0.199]	−0.041 [−0.083, 0.132]
Mental health (δ_2)	**−0.007** [−0.010, −0.004]	**−0.008** [−0.012, −0.005]	**−0.006** [−0.011, −0.003]	**0.931** [0.843, 1.048]	**1.017** [0.801, 1.064]	**1.041** [0.868, 1.083]
Complementarity (ρ)	**0.763** [0.616, 0.892]	**0.572** [0.339, 0.772]	**0.990** [0.828, 1.153]	0.303 [−0.241, 0.634]	0.617 [−0.112, 1.005]	0.413 [−0.181, 0.611]
Elasticity of substitution	**4.214** [2.604, 9.224]	**2.335** [1.513, 4.387]	6.764 [−206.381, 216.522]	**1.425** [0.790, 2.710]	2.536 [−2.794, 14.189]	**1.703** [0.837, 2.557]
Complementarity – Lagged hum. capital (ρ^L)	−0.172 [−0.343, 0.003]	−0.094 [−0.286, 0.110]	−0.217 [−0.578, 0.118]	0.561 [−4.319, 1.038]	−1.061 [−10.729, 0.992]	−2.635 [−3.849, 0.823]
Elasticity of substitution – Lagged hum. capital	**0.853** [0.742, 1.002]	**0.914** [0.776, 1.115]	**0.820** [0.631, 1.115]	2.183 [−11.329, 18.305]	0.474 [−0.743, 8.181]	**0.274** [0.198, 4.925]
TFP (A)	**1.332** [1.278, 1.400]	**1.282** [1.206, 1.370]	**1.456** [1.371, 1.574]	**0.833** [0.614, 1.235]	**1.044** [0.518, 1.292]	**1.055** [0.605, 1.262]
Diet quality control function (φ^D)	−0.002 [−0.005, 0.001]	**−0.004** [−0.008, −0.000]	0.000 [−0.003, 0.004]	−0.004 [−0.043, 0.036]	−0.013 [−0.066, 0.043]	−0.013 [−0.059, 0.036]
Physical activity control function (φ^E)	**−0.092** [−0.108, −0.078]	**−0.080** [−0.103, −0.061]	**−0.108** [−0.135, −0.089]	0.100 [−0.049, 0.226]	0.023 [−0.055, 0.281]	−0.019 [−0.091, 0.213]
Estimator convergence rate	100%	100%	99.9%	100%	99.5%	100%

Notes: The table presents estimates of the parameters of a CES production function. It is estimated using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged in to the second stage CES equation. The CES function is estimated by nonlinear least squares using the Levenberg-Marquardt algorithm (Elzhov et al., 2016). The control function procedure is bootstrapped with 1000 repetitions: medians from the bootstrapped distribution are on top, and 95% confidence intervals are in brackets. *N*: 4,802 (all), 2,447 (males), 2,355 (females). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

4.8 Figures

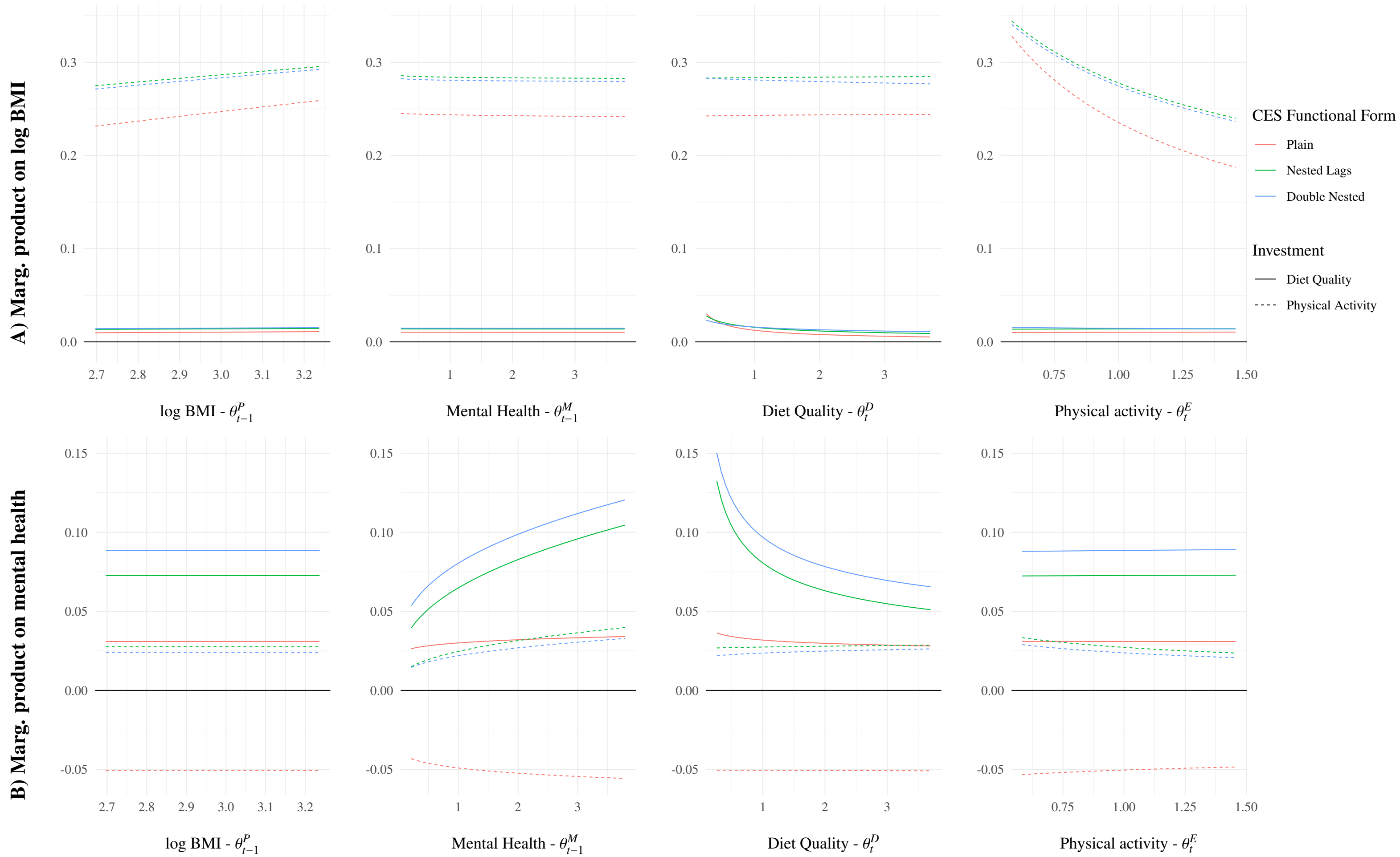

Figure 4.1: Marginal product of investments – males

Notes: The graphs show marginal products of diet quality and physical activity on log BMI (Panel A, top) and mental health (Panel B, bottom). The products are computed using the estimated parameters from the control function CES estimation in Table 4.3, Table 4.A6, and Table 4.A8. Each plot shows how marginal products vary between the 5th and 95th percentile of the four inputs of the CES function.

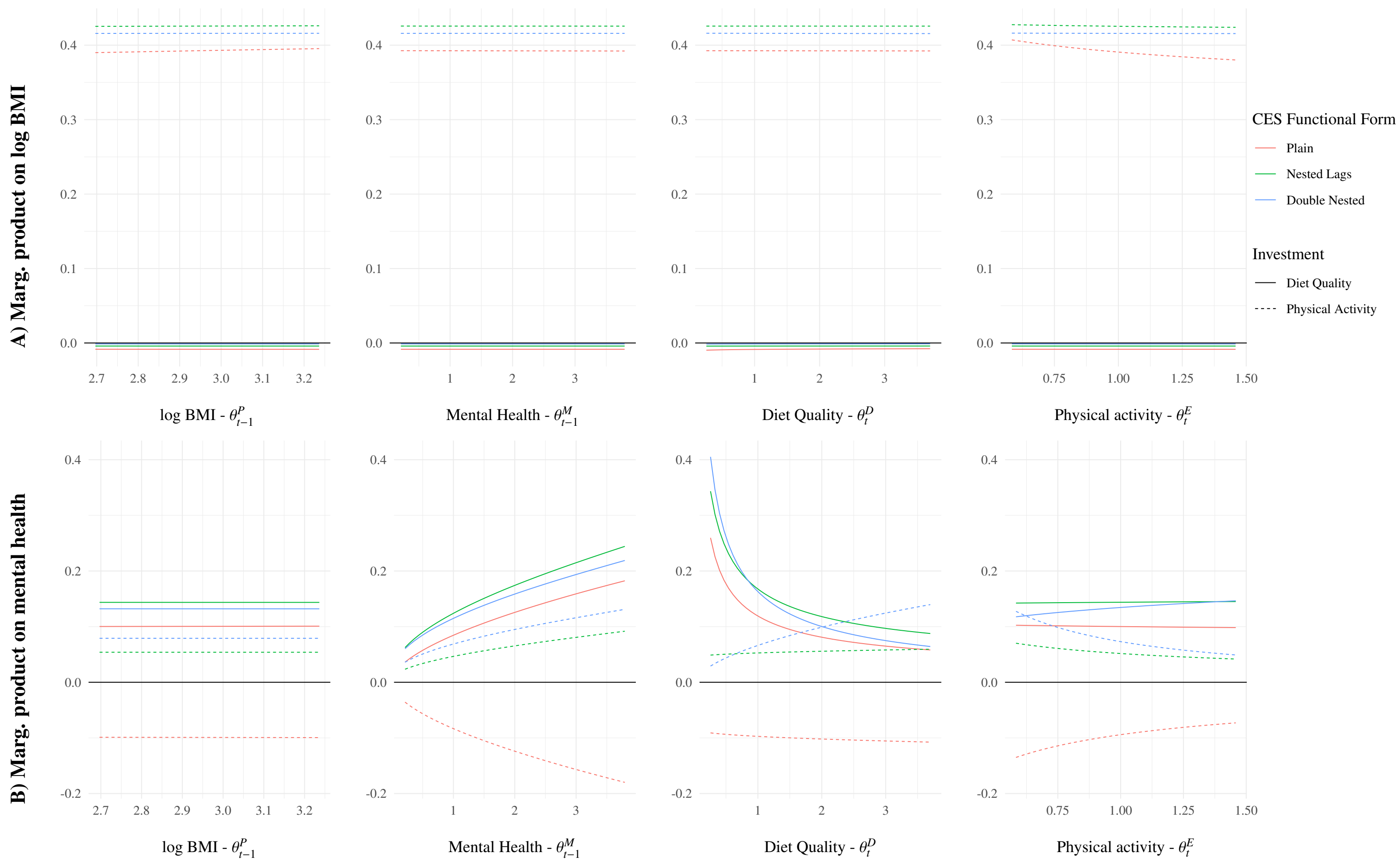

Figure 4.2: Marginal product of investments – females

Notes: The graphs show marginal products of diet quality and physical activity on log BMI (Panel A, top) and mental health (Panel B, bottom). The products are computed using the estimated parameters from the control function CES estimation in Table 4.3, Table 4.A6, and Table 4.A8. Each plot shows how marginal products vary between the 5th and 95th percentile of the four inputs of the CES function.

4.9 Appendix tables

Table 4.A1: MCS variable description

Variable Group	MCS Survey	Variable	Note
Father's occupational social class	First (9 months)	White collar (I-IIINM)^d Blue collar (IIIM-V-other)^d No father figure^d	Based on father's Registrar General Social Class classification of occupations. White collar includes I (Professional), II (Managerial/technical), IIINM (Skilled non-manual). Blue collar includes IIIM (Skilled manual), IV (Partly skilled), V (Unskilled), Other, Unemployed, and Armed forces. No father figure is a dummy for children whose father does not live in the same household.
Maternal education	First (9 months)	National Vocational Qualification (NVQ) Equivalents^d	Categorical, with levels: *No qual. - NVQ 1* (no qualifications, or equivalent to GCSE grades D-G); *NVQ 2* (equivalent to GCSE grades A*-C); *NVQ 3* (equivalent to A/AS levels); *NVQ 4-5* (higher education qualification); *Other qualification* (e.g foreign degrees).
Family background	First (9 months)	Mother's age at birth (years) Child nonwhite ethnicity^d Single-parent household	All variables are self-reported by the mother at birth.
Family income	First (9 months)	Family income below 60% of median	Family income is equivalised using the OECD scale, and then compared to the national median family income.
Initial Health	First (9 months)	Birth weight	Reported by mother, consulting child's personal health record ("red book")
Rural	Fifth (11y)	Resides in rural area	Derived from location at interview
Physical Health (θ_t^P)	Fifth (11y) Sixth (14y)	Body Mass Index (BMI) Overweight (incl. obese)^d Body fat percentage	Height, weight, and body fat are measured by trained interviewers using scales with bioelectrical impedance capability (Connelly and Platt, 2014). Overweight is calculated with reference to IOTF cutoffs by gender and age (Cole and Lobstein, 2012), with age measured to the nearest tenth of a year.
Mental Health (θ_t^M)	Fifth (11y) Sixth (14y)	SDQ Total Difficulties Items^d	Twenty items from the Strength and Difficulties Questionnaire (Goodman, 1997), as reported by the mother.
Diet quality (θ_t^D)	Sixth (14y)	Six questions from the child interview^d	The questions are: (i) *How often do you eat breakfast over a week?* (ii) *How often do you eat at least 2 portions of fruit per day?* (iii) *How often do you eat at least 2 portions of vegetables including salad, fresh, frozen or tinned vegetables per day?* (iv) *Which type of bread do you normally eat? (Wholemeal vs white)* (v) *How often, if at all, do you drink sugary drinks like regular cola or squash?* (vi) *How often, if at all, do you eat fast food such as McDonalds, Burger King, KFC or other fast food like that?*
Physical activity / exercise (θ_t^E)	Sixth (14y)	Question from the child interview^d	The phrasing of the question is: *On how many days in the last week did you do a total of at least an hour of moderate to vigorous physical activity? By moderate to vigorous we mean any physical activity that makes you get warmer, breathe harder and makes your heart beat faster, e.g. riding a bike, running, playing football, swimming, dancing, etc.*

Notes: Variables denoted by ^{d} are binary or categorical.

Table 4.A2: Instrumental variables

Panel A: Components of price indices

Index	Item Code	Item Label	Index	Item Code	Item Label
White bread	210102	Large Loaf-White-Unsliced-800G	Eggs	211602	Eggs-Medium-Per Doz Or 2 X 6
	210111	White Sliced Loaf Branded 750G		211603	Eggs-Large-Per Doz Or 2 X 6
Fizzy drinks	212011	Cola Flavd Drink 1.75-2 Lt Btl	Cereal	210213	Breakfast Cereal 1
	212015	Fizzy Energy Drink 250-500Ml		210214	Breakfast Cereal 2
	212017	Fizzy Bottled Drink 500Ml		210216	Hot Oat Cereal
	220320	Takeaway Soft Drink			
Vegetables	212504	Fresh Veg-Cauliflower-Each	Fresh fruit	212709	Oranges-Class 1-Each
	212510	Fresh Veg-Cucumber-Whole		212710	Avocado Pear-Each
	212511	Fresh Veg-Lettuce-Iceberg-Each		212712	Kiwi Fruit-Each
	212515	Fresh Veg-Tomatoes-Per Kg		212715	Grapefruit-Each
	212516	Fresh Veg-Cabbage-Whole-Per Kg		212716	Apples-Cooking-Per Kg
	212518	Fresh Veg-Carrots-Per Kg		212717	Apples-Dessert-Per Kg
	212519	Fresh Veg-Onions-Per Kg		212718	Pears-Dessert-Per Kg
	212520	Fresh Veg-Mushrooms-Per Kg		212719	Bananas-Per Kg
	212527	Pre-Packed Salad 100-250G		212720	Strawberries Per Kg Or Punnet
	212530	Fresh Veg-Broccoli-Per Kg		212722	Grapes-Per Kg
	212531	Fresh Veg-Courgette-Per Kg		212725	Small Type Oranges Per Pack/Kg
	212532	Fresh Veg-Peppers -Loose Or Kg		212726	Plums Per Kg/Pack
	212601	Canned Tomatoes 390-400G		212727	Peach/Nectarine Each
	212603	Baked Beans, 400-420G Tin		212728	Pineapple Each
	212608	Canned Sweetcorn 198-340G		212729	Blueberries Punnet Per Kg
	212609	Frozen Garden Peas 800G-1Kg		212730	Fruit Fresh Snacking 150-350G
Sweets	212211	Chocolate 4	Takeaway	220301	Fish & Chips Takeaway
	212218	Carton/Box Of Chocs 150-400Gm		220303	Sandwich-Take-Away (Cold)
	212219	Chocolate 8		220316	Pizza Takeaway Or Delivered
	212222	Chocolate 10		220317	Pasty/Savoury Pie - Takeaway
	212223	Brand Choc Sweets 100-185G Bag		220318	Indian Takeaway
	212224	Sweets, Not Choc, 150-250G Bag		220319	Chinese Takeaway
Potato crisps	212402	Potato Crisps-25G/40G		220322	Burger In Bun- Takeaway
	212404	Potato Crisps-Multi-Pack		220323	Kebab- Takeaway
	220310	Potato Crisps-Individual Pack			

Panel B: Weather variables

Variable	Note	Unit
Total precipitation	Total precipitation amount over the calendar month	Millimeters
Days rain	Count of days with more than 1mm precipitation	Days
Temperature	Average of daily mean air temperature over the calendar month	Degrees Celsius
Subshine	Duration of bright sunshine during the month	Hours

Notes: Panel A of the table shows the items from the ONS CPI price survey that underlie the price indices used as instruments for diet quality. These vary by month and Government Office Region. Panel B shows the available information in the weather data, used as instruments for physical activity. These vary by month and Output Area.

Table 4.A3: Summary Statistics

	All	Males	Females
Birth weight (kg)	3.39 (0.57)	3.44 (0.58)	3.33 (0.55)
Mother age at birth	29.44 (5.61)	29.43 (5.51)	29.45 (5.73)
Country of birth			
Born in England	0.76	0.75	0.76
Born in NI	0.00	0.00	0.00
Born in Scotland	0.01	0.01	0.00
Born in Wales	0.24	0.24	0.24
Non-white ethnicity	0.13	0.13	0.13
Family income under .6 median	0.23	0.23	0.24
Father occ. social class at birth			
Blue collar	0.44	0.43	0.45
No father fig.	0.10	0.10	0.11
White collar	0.45	0.46	0.44
Maternal education level			
No qual / NVQ 1	0.16	0.15	0.16
GCSE and equiv.	0.28	0.28	0.28
A-Level and equiv.	0.14	0.14	0.14
Higher education	0.40	0.41	0.39
Other qual.	0.02	0.02	0.02
11y interview			
Age	10.66 (0.49)	10.66 (0.49)	10.66 (0.48)
BMI	18.98 (3.35)	18.84 (3.23)	19.14 (3.46)
SDQ Tot Diff score	7.15 (5.49)	7.68 (5.67)	6.61 (5.24)
Interv. in Winter	0.38	0.38	0.38
Interv. in Spring	0.45	0.44	0.46
Interv. in Summer	0.15	0.15	0.14
Interv. in Autumn	0.02	0.03	0.02
14y interview			
Age	13.75 (0.45)	13.75 (0.46)	13.76 (0.45)
BMI	21.29 (3.99)	20.79 (3.84)	21.81 (4.08)
Num. siblings in HH	1.43 (1.02)	1.46 (1.01)	1.40 (1.03)
SDQ Tot Diff score	7.59 (5.71)	7.86 (5.85)	7.31 (5.54)
Interv. in Winter	0.22	0.23	0.22
Interv. in Spring	0.43	0.42	0.43
Interv. in Summer	0.29	0.28	0.29
Interv. in Autumn	0.06	0.06	0.06

Notes: The table shows summary statistics for the sample used in estimating human capital production functions. Variable definitions are in Table 4.A1.

Table 4.A4: Diet quality EFA

	All	Males	Females
Vegetables	0.698	0.698	0.700
Fruit	0.649	0.639	0.662
Fast food	0.539	0.478	0.590
Days breakfast	0.403	0.394	0.469
Wholemeal bread	0.436	0.415	0.444
Milk fat	0.170	0.175	0.160
Sugary drinks	0.214	0.159	0.263
Artificially sweetened drinks	0.493	0.445	0.532

Notes: The table shows estimated factor loadings from an exploratory factor analysis (EFA) on the diet quality data in the MCS age 14 survey. First, the polychoric correlation matrix of the data is estimated, and then Minimum Residual EFA is applied. It uses **R** packages `polycor` (Fox, 2016) and `psych` (Revelle, 2018).

Table 4.A5: Measurement system

Factor	Measure	All Loading (λ)	All % Signal	Males Loading (λ)	Males % Signal	Females Loading (λ)	Females % Signal
Mental health (14y)	SDQ Adults	0.460	55.9	0.456	55.8	0.473	56.3
Mental health (14y)	SDQ Attentive	0.725	67.8	0.717	67.3	0.714	67.1
Mental health (14y)	SDQ Bullied	0.602	61.1	0.607	61.3	0.615	61.6
Mental health (14y)	SDQ Clingy	0.574	59.8	0.586	60.4	0.591	60.6
Mental health (14y)	SDQ Distract	0.758	70.2	0.753	69.8	0.746	69.3
Mental health (14y)	SDQ Fears	0.599	60.9	0.596	60.8	0.631	62.4
Mental health (14y)	SDQ Fidgety	0.740	68.9	0.740	68.8	0.720	67.5
Mental health (14y)	SDQ Fights	0.688	65.5	0.724	67.8	0.637	62.7
Mental health (14y)	SDQ Goodfriend	0.528	58.1	0.550	58.9	0.493	56.9
Mental health (14y)	SDQ Lies	0.667	64.3	0.673	64.6	0.648	63.3
Mental health (14y)	SDQ Obedient	0.537	58.4	0.554	59.1	0.514	57.6
Mental health (14y)	SDQ Popular	0.643	63.0	0.660	63.9	0.626	62.2
Mental health (14y)	SDQ Reflective	0.621	62.0	0.633	62.5	0.593	60.7
Mental health (14y)	SDQ Restless	0.649	63.4	0.671	64.5	0.595	60.8
Mental health (14y)	SDQ Solitary	0.413	54.7	0.412	54.6	0.415	54.7
Mental health (14y)	SDQ Somatic	0.403	54.4	0.421	54.9	0.446	55.5
Mental health (14y)	SDQ Steals	0.685	65.3	0.709	66.8	0.632	62.5
Mental health (14y)	SDQ Tempers	0.641	62.9	0.645	63.1	0.651	63.4
Mental health (14y)	SDQ Unhappy	0.609	61.4	0.649	63.3	0.616	61.7
Mental health (14y)	SDQ Worries	0.590	60.5	0.605	61.2	0.615	61.7
Mental health (11y)	SDQ Adults	0.476	56.4	0.456	55.8	0.502	57.2
Mental health (11y)	SDQ Attentive	0.685	65.3	0.682	65.2	0.663	64.1
Mental health (11y)	SDQ Bullied	0.612	61.5	0.623	62.1	0.608	61.3
Mental health (11y)	SDQ Clingy	0.521	57.8	0.515	57.6	0.551	58.9
Mental health (11y)	SDQ Distract	0.744	69.1	0.733	68.4	0.739	68.8
Mental health (11y)	SDQ Fears	0.591	60.6	0.598	60.9	0.600	61.0
Mental health (11y)	SDQ Fidgety	0.720	67.5	0.729	68.1	0.691	65.7
Mental health (11y)	SDQ Fights	0.704	66.5	0.722	67.6	0.674	64.7
Mental health (11y)	SDQ Goodfriend	0.477	56.4	0.481	56.5	0.471	56.2
Mental health (11y)	SDQ Lies	0.614	61.6	0.588	60.4	0.637	62.7
Mental health (11y)	SDQ Obedient	0.573	59.8	0.588	60.5	0.550	58.9
Mental health (11y)	SDQ Popular	0.640	62.9	0.649	63.3	0.634	62.6
Mental health (11y)	SDQ Reflective	0.623	62.0	0.638	62.8	0.585	60.3
Mental health (11y)	SDQ Restless	0.713	67.0	0.726	67.9	0.676	64.8
Mental health (11y)	SDQ Solitary	0.437	55.3	0.431	55.1	0.441	55.4
Mental health (11y)	SDQ Somatic	0.391	54.1	0.409	54.6	0.427	55.0
Mental health (11y)	SDQ Steals	0.604	61.2	0.682	65.1	0.513	57.6
Mental health (11y)	SDQ Tempers	0.661	64.0	0.671	64.5	0.655	63.7
Mental health (11y)	SDQ Unhappy	0.673	64.7	0.670	64.5	0.688	65.5
Mental health (11y)	SDQ Worries	0.590	60.5	0.588	60.5	0.606	61.2
Diet quality (14y)	Days breakfast	0.476	56.4	0.468	56.2	0.576	59.9
Diet quality (14y)	Wholemeal bread	0.440	55.4	0.391	54.1	0.477	56.4
Diet quality (14y)	Fruit	0.664	64.2	0.643	63.0	0.689	65.5
Diet quality (14y)	Vegetables	0.735	68.5	0.741	68.9	0.723	67.7
Diet quality (14y)	Fast food	0.521	57.8	0.492	56.9	0.529	58.1
Diet quality (14y)	Artificially sweetened drinks	0.465	56.1	0.426	55.0	0.476	56.4

Notes: The table shows estimated factor loadings and signal ratios from the measurement system in (4.4.1). The columns *% signal* show the percentage of the variance of each measure which can be explained by the latent factor. This value is obtained as $100 \cdot 1/\mathrm{Var}(\epsilon)$, since the variance of the latent factors is normalised to 1. All items are recoded so that higher values correspond to better mental health or diet quality, hence the positive factor loadings.

Table 4.A6: CES production functions – Plain

$$\theta_{i,t}^{k} = A_t^k \left[\gamma_1^k {\theta_{i,t-1}^{P}}^{\rho_k} + \gamma_2^k {\theta_{i,t-1}^{M}}^{\rho_k} + \gamma_3^k (\theta_{i,t}^{D})^{\rho_k} + \gamma_4^k (\theta_{i,t}^{E})^{\rho_k}\right]^{\frac{1}{\rho_k}} \cdot \exp\left\{\boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varphi_k^D \hat{v}_{i,t}^{D} + \varphi_k^E \hat{v}_{i,t}^{E} + \varepsilon_{i,t}^{k}\right\} \qquad \text{with } \sum_{l=1}^{4}\gamma_l^k = 1$$

	log BMI – θ_t^P			**Mental health** – θ_t^M		
	All	Males	Females	All	Males	Females
log BMI (γ_1)	**0.853**	**0.905**	**0.756**	0.034	**0.037**	0.040
	[0.810, 0.893]	[0.854, 0.939]	[0.684, 0.823]	[−0.002, 0.144]	[0.006, 0.293]	[−0.053, 0.180]
Mental health (γ_2)	**−0.010**	**−0.010**	**−0.009**	**0.958**	**0.983**	**0.935**
	[−0.013, −0.007]	[−0.014, −0.007]	[−0.013, −0.004]	[0.915, 0.997]	[0.926, 1.055]	[0.881, 0.992]
Diet quality (γ_3)	0.002	0.005	−0.006	**0.065**	0.035	**0.112**
	[−0.005, 0.007]	[−0.000, 0.011]	[−0.012, 0.002]	[0.009, 0.126]	[−0.005, 0.113]	[0.006, 0.182]
Physical activity (γ_4)	**0.156**	**0.100**	**0.258**	−0.048	−0.058	−0.079
	[0.115, 0.201]	[0.066, 0.151]	[0.190, 0.331]	[−0.222, 0.029]	[−0.340, 0.028]	[−0.289, 0.051]
Complementarity (ρ)	**0.585**	**0.342**	**0.916**	**0.691**	**0.900**	**0.374**
	[0.398, 0.776]	[0.135, 0.570]	[0.689, 1.125]	[0.196, 1.187]	[0.108, 1.329]	[0.043, 1.245]
Elasticity of substitution	**2.412**	**1.520**	6.703	1.840	1.595	1.411
	[1.662, 4.471]	[1.156, 2.324]	[−124.986, 110.715]	[−45.036, 65.492]	[−63.072, 79.050]	[−18.486, 29.363]
TFP (A)	**1.285**	**1.232**	**1.413**	**0.927**	**0.894**	**0.849**
	[1.230, 1.340]	[1.173, 1.310]	[1.326, 1.521]	[0.730, 1.107]	[0.489, 1.110]	[0.600, 1.186]
Diet quality control function (φ^D)	−0.001	−0.003	0.002	0.012	0.007	0.006
	[−0.004, 0.002]	[−0.006, 0.000]	[−0.001, 0.005]	[−0.026, 0.044]	[−0.045, 0.038]	[−0.041, 0.064]
Physical activity control function (φ^E)	**−0.084**	**−0.072**	**−0.102**	0.035	0.052	0.063
	[−0.097, −0.071]	[−0.092, −0.056]	[−0.124, −0.083]	[−0.033, 0.173]	[−0.030, 0.312]	[−0.055, 0.229]
Estimator convergence rate	100%	99.2%	99.9%	99.9%	99.8%	99.9%

Notes: The table presents estimates of the parameters of a CES production function. It is estimated using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged in to the second stage CES equation. The CES function is estimated by nonlinear least squares using the Levenberg-Marquardt algorithm (Elzhov et al., 2016). The control function procedure is bootstrapped with 1000 repetitions: medians from the bootstrapped distribution are on top, and 95% confidence intervals are in brackets. N: 4,802 (all), 2,447 (males), 2,355 (females). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

Table 4.A7: CES production functions – Nested investments

$$\theta^k_{i,t} = A^k_t \left[\gamma^k_1 {\theta^P_{i,t-1}}^{\rho_k} + \gamma^k_2 {\theta^M_{i,t-1}}^{\rho_k} + \gamma^k_3 \left(\delta^k_1 \left(\theta^D_{i,t}\right)^{\rho^I_k} + \delta^k_2 \left(\theta^E_{i,t}\right)^{\rho^I_k} \right)^{\rho_k/\rho^I_k} \right]^{\frac{1}{\rho_k}} \cdot \exp\left\{ \boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varphi^D_k \hat{v}^D_{i,t} + \varphi^E_k \hat{v}^E_{i,t} + \varepsilon^k_{i,t} \right\} \qquad \text{with } \gamma^k_1 + \gamma^k_2 + \gamma^k_3 = 1,\ \delta^k_1 + \delta^k_2 = 1.$$

	log BMI – θ^P_t			Mental health – θ^M_t		
	All	Males	Females	All	Males	Females
log BMI (γ_1)	**0.854** [0.808, 0.893]	**0.909** [0.860, 0.943]	**0.761** [0.690, 0.827]	0.015 [−0.020, 0.031]	0.017 [−0.013, 0.046]	0.006 [−0.056, 0.035]
Mental health (γ_2)	**−0.010** [−0.013, −0.008]	**−0.010** [−0.014, −0.007]	**−0.009** [−0.013, −0.005]	**0.929** [0.898, 0.960]	**0.951** [0.913, 0.997]	**0.900** [0.854, 0.944]
Investment composite (γ_3)	**0.156** [0.118, 0.202]	**0.102** [0.065, 0.151]	**0.248** [0.184, 0.317]	**0.056** [0.012, 0.115]	0.030 [−0.047, 0.099]	**0.098** [0.032, 0.179]
Diet quality (δ_1)	0.016 [−0.019, 0.056]	0.056 [−0.006, 0.119]	0.000 [−0.048, 0.019]	**1.004** [0.049, 1.577]	**1.072** [0.000, 2.234]	**1.000** [0.010, 2.670]
Physical activity (δ_2)	**0.984** [0.944, 1.019]	**0.944** [0.881, 1.006]	**1.000** [0.981, 1.048]	−0.004 [−0.577, 0.951]	−0.072 [−1.234, 1.000]	0.000 [−1.670, 0.990]
Complementarity (ρ)	**0.583** [0.400, 0.769]	**0.308** [0.079, 0.556]	**0.895** [0.678, 1.091]	**0.812** [0.340, 1.166]	**0.981** [0.386, 1.333]	**0.549** [0.108, 1.246]
Elasticity of substitution	**2.391** [1.647, 4.225]	**1.444** [1.082, 2.212]	6.962 [−56.469, 83.891]	2.822 [−63.968, 70.433]	1.955 [−72.750, 76.512]	1.693 [−18.178, 41.035]
Complementarity – inputs (ρ^I)	0.360 [−1.242, 1.823]	0.769 [−1.014, 2.567]	0.083 [−3.465, 1.915]	0.313 [−31.910, 1.617]	0.562 [−14.793, 24.392]	−0.918 [−26.079, 1.036]
Elasticity of substitution – inputs	1.245 [−6.849, 10.618]	2.325 [−24.775, 29.926]	0.507 [−23.436, 14.990]	0.954 [−0.421, 6.691]	0.994 [−9.723, 12.446]	0.479 [−0.103, 6.238]
TFP (A)	**1.282** [1.232, 1.343]	**1.228** [1.169, 1.308]	**1.411** [1.325, 1.518]	**0.973** [0.856, 1.152]	**0.961** [0.775, 1.186]	**0.975** [0.789, 1.234]
Diet quality control function (φ^D)	−0.001 [−0.004, 0.002]	−0.004 [−0.007, 0.000]	0.000 [−0.002, 0.004]	0.020 [−0.020, 0.045]	0.010 [−0.042, 0.037]	0.022 [−0.037, 0.067]
Physical activity control function (φ^E)	**−0.083** [−0.097, −0.070]	**−0.071** [−0.092, −0.054]	**−0.100** [−0.120, −0.080]	0.006 [−0.042, 0.062]	0.026 [−0.032, 0.092]	−0.001 [−0.072, 0.085]
Estimator convergence rate	98.5%	99.3%	96.2%	79.2%	76.7%	64.6%

Notes: The table presents estimates of the parameters of a CES production function. It is estimated using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged in to the second stage CES equation. The CES function is estimated by nonlinear least squares using the Levenberg-Marquardt algorithm (Elzhov et al., 2016). The control function procedure is bootstrapped with 1000 repetitions: medians from the bootstrapped distribution are on top, and 95% confidence intervals are in brackets. *N*: 4,802 (all), 2,447 (males), 2,355 (females). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

Table 4.A8: CES Health production functions – Double Nesting

$$\theta^k_{i,t} = A^k_t \left[\gamma^k_1 \left(\delta^k_1 \theta^P_{i,t-1}{}^{\rho^L_k} + \delta^k_2 \theta^M_{i,t-1}{}^{\rho^L_k} \right)^{\rho_k/\rho^L_k} + \gamma^k_2 \left(\delta^k_3 \left(\theta^D_{i,t}\right)^{\rho^I_k} + \delta^k_4 \left(\theta^E_{i,t}\right)^{\rho^I_k} \right)^{\rho_k/\rho^I_k} \right]^{\frac{1}{\rho_k}} \cdot \exp\left\{ \boldsymbol{X}_{i,t}\boldsymbol{\xi} + \varphi^D_k \hat{v}^D_{i,t} + \varphi^E_k \hat{v}^E_{i,t} + \varepsilon^k_{i,t} \right\} \quad \text{with} \quad \gamma^k_1 + \gamma^k_2 = 1, \delta^k_1 + \delta^k_2 = 1, \delta^k_3 + \delta^k_4 = 1.$$

	log BMI – θ^P_t			Mental health – θ^M_t		
	All	Males	Females	All	Males	Females
Lagged hum. capital composite (γ_1)	**0.791** [0.745, 0.832]	**0.851** [0.790, 0.907]	**0.714** [0.645, 0.780]	**0.867** [0.844, 0.890]	**0.895** [0.853, 0.976]	**0.821** [0.786, 0.854]
Investment composite (γ_2)	**0.209** [0.168, 0.255]	**0.149** [0.093, 0.210]	**0.286** [0.220, 0.355]	**0.133** [0.110, 0.156]	**0.105** [0.024, 0.147]	**0.179** [0.146, 0.214]
log BMI (δ_1)	**1.007** [1.004, 1.010]	**1.008** [1.004, 1.012]	**1.006** [1.000, 1.011]	**−0.016** [−0.048, −0.004]	−0.023 [−0.068, 0.025]	**−0.043** [−0.088, −0.018]
Mental health (δ_2)	**−0.007** [−0.010, −0.004]	**−0.008** [−0.012, −0.004]	−0.006 [−0.011, 0.000]	**1.016** [1.004, 1.048]	**1.023** [0.975, 1.068]	**1.043** [1.018, 1.088]
Diet quality (δ_3)	0.025 [−0.009, 0.055]	0.054 [−0.006, 0.110]	−0.003 [−0.069, 0.030]	**0.823** [0.496, 1.125]	**0.806** [0.244, 1.427]	**0.730** [0.416, 1.034]
Physical activity (δ_4)	**0.975** [0.945, 1.009]	**0.946** [0.890, 1.006]	**1.003** [0.970, 1.069]	0.177 [−0.125, 0.504]	0.194 [−0.427, 0.756]	0.270 [−0.034, 0.584]
Complementarity (ρ)	**0.781** [0.642, 0.916]	**0.555** [0.308, 0.767]	**0.998** [0.827, 1.185]	**0.528** [0.355, 0.707]	**0.675** [0.347, 1.026]	**0.442** [0.238, 0.631]
Elasticity of substitution	**4.559** [2.794, 11.853]	**2.248** [1.445, 4.290]	5.131 [−187.941, 210.338]	**2.119** [1.551, 3.416]	2.961 [−9.431, 14.280]	**1.794** [1.312, 2.707]
Complementarity – Lagged hum. capital (ρ^L)	−0.163 [−0.339, 0.054]	−0.085 [−0.303, 0.106]	−0.184 [−0.522, 18.117]	**−3.337** [−5.133, −1.787]	−1.322 [−9.424, 1.166]	**−2.727** [−3.839, −1.675]
Elasticity of substitution – Lagged hum. capital	**0.857** [0.724, 1.005]	**0.921** [0.764, 1.109]	0.808 [−0.394, 1.073]	**0.230** [0.162, 0.356]	0.396 [−8.736, 1.614]	**0.268** [0.207, 0.374]
Complementarity – inputs (ρ^I)	0.217 [−0.883, 1.490]	0.723 [−0.148, 1.869]	0.904 [−2.615, 2.134]	−0.082 [−9.734, 1.421]	0.621 [−9.185, 26.901]	−0.281 [−9.127, 0.982]
Elasticity of substitution – inputs	1.207 [−2.281, 6.375]	2.507 [−26.287, 39.187]	0.329 [−27.314, 20.067]	0.831 [−4.683, 7.928]	0.632 [−19.076, 12.787]	**0.751** [0.024, 4.359]
TFP (A)	**1.335** [1.279, 1.407]	**1.278** [1.203, 1.371]	**1.451** [1.346, 1.564]	**1.104** [0.969, 1.244]	**1.085** [0.887, 1.331]	**1.069** [0.899, 1.302]
Diet quality control function (φ^D)	−0.002 [−0.005, 0.001]	−0.004 [−0.008, 0.000]	−0.000 [−0.003, 0.006]	−0.012 [−0.046, 0.020]	−0.017 [−0.074, 0.022]	−0.001 [−0.047, 0.041]
Physical activity control function (φ^E)	**−0.091** [−0.108, −0.077]	**−0.079** [−0.105, −0.060]	**−0.107** [−0.131, −0.082]	−0.017 [−0.061, 0.029]	0.011 [−0.059, 0.073]	−0.040 [−0.105, 0.023]
Estimator convergence rate	99.7%	99.2%	98.4%	88.4%	82.6%	89.4%

Notes: The table presents estimates of the parameters of a CES production function. It is estimated using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged in to the second stage CES equation. The CES function is estimated by nonlinear least squares using the Levenberg-Marquardt algorithm (Elzhov et al., 2016). The control function procedure is bootstrapped with 1000 repetitions: medians from the bootstrapped distribution are on top, and 95% confidence intervals are in brackets. *N*: 4,802 (all), 2,447 (males), 2,355 (females). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

Table 4.A9: Exogenous production functions

Panel A: Exogenous vs Endogenous nested lags CES specification

	log BMI – θ_t^P		**Mental health** – θ_t^M	
	Exog	Endog	Exog	Endog
Lagged hum. capital composite (γ_1)	$\mathbf{0.998}$	$\mathbf{0.794}$	$\mathbf{0.955}$	$\mathbf{1.051}$
	[0.993, 1.003]	[0.747, 0.835]	[0.923, 0.986]	[0.846, 1.171]
Diet quality (γ_2)	0.000	0.003	$\mathbf{0.081}$	$\mathbf{0.085}$
	[−0.001, 0.002]	[−0.003, 0.009]	[0.067, 0.094]	[0.012, 0.149]
Physical activity (γ_3)	$\mathbf{0.002}$	$\mathbf{0.202}$	$\mathbf{-0.035}$	−0.115
	[0.994, 1.002]	[0.159, 0.252]	[1.003, 1.068]	[−0.245, 0.050]
log BMI (δ_1)	$\mathbf{1.001}$	$\mathbf{1.007}$	0.014	0.069
	[1.000, 1.002]	[1.004, 1.010]	[−0.006, 0.034]	[−0.048, 0.157]
Mental health (δ_2)	$\mathbf{-0.001}$	$\mathbf{-0.007}$	0.986	$\mathbf{0.931}$
	[1.000, 1.002]	[−0.010, −0.004]	[−0.006, 0.034]	[0.843, 1.048]
Complementarity (ρ)	−0.495	$\mathbf{0.763}$	$\mathbf{0.415}$	0.303
	[−2.050, 1.060]	[0.616, 0.892]	[0.192, 0.638]	[−0.241, 0.634]
Complementarity – Lagged HC (ρ^L)	−0.505	−0.172	$\mathbf{1.327}$	0.561
	[−1.035, 0.025]	[−0.343, 0.003]	[0.597, 2.057]	[−4.319, 1.038]
TFP (A)	$\mathbf{1.035}$	$\mathbf{1.332}$	$\mathbf{1.024}$	$\mathbf{0.833}$
	[1.025, 1.045]	[1.278, 1.400]	[0.938, 1.111]	[0.614, 1.235]

Panel B: Exogenous vs Endogenous linear specification

	log BMI – θ_t^P		**Mental health** – θ_t^M	
	Exog	Endog	Exog	Endog
log BMI (11) – θ_{t-1}^P	$\mathbf{0.874}$	$\mathbf{0.907}$	0.070	0.022
	[0.859, 0.890]	[0.873, 0.964]	[−0.012, 0.152]	[−0.174, 0.192]
Mental health (11) – θ_{t-1}^M	$\mathbf{-0.004}$	$\mathbf{-0.011}$	$\mathbf{0.924}$	$\mathbf{0.929}$
	[−0.006, −0.002]	[−0.020, −0.004]	[0.913, 0.936]	[0.893, 0.965]
Diet quality – θ_t^D	−0.001	0.016	$\mathbf{0.070}$	0.083
	[−0.003, 0.001]	[−0.011, 0.045]	[0.058, 0.083]	[−0.050, 0.214]
Physical activity – θ_t^E	−0.008	$\mathbf{0.178}$	0.003	−0.270
	[−0.016, 0.000]	[0.013, 0.487]	[−0.039, 0.045]	[−1.274, 0.592]

Notes: The table presents estimates of the parameters of CES and linear production functions for the entire sample of MCS children. Estimates in the *Exog* columns are obtained without any correction for the potential endogeneity of diet quality and physical activity. Estimates in the *Endog* columns are control function estimates that correct for the endogeneity of diet quality and physical activity. Panel A presents estimates from a nested lags CES specification using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged into the second stage CES equation. The CES function is estimated by nonlinear least squares using the Levenberg-Marquardt algorithm (Elzhov et al., 2016). Panel B presents linear production function estimates. These are also obtained using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged into the second stage linear equation. In both cases, bootstrap 95% confidence intervals with 1000 repetitions are in brackets. N: 4,802 (all). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

Table 4.A10: Linear production functions with alternative measures of physical health

Panel A: Linear production function – Body fat percentage

	log Body fat % – θ_t^P		**Mental health – θ_t^M**	
	Exog	Endog	Exog	Endog
log Body fat % (11) – θ_{t-1}^P	$\mathbf{0.778}$ [0.756, 0.800]	$\mathbf{0.782}$ [0.727, 0.847]	$\mathbf{0.057}$ [0.016, 0.098]	0.022 [−0.089, 0.122]
Mental health (11) – θ_{t-1}^M	$\mathbf{-0.007}$ [−0.013, −0.001]	0.017 [−0.002, 0.040]	$\mathbf{0.926}$ [0.914, 0.938]	$\mathbf{0.932}$ [0.897, 0.967]
Diet quality – θ_t^D	−0.004 [−0.011, 0.003]	$\mathbf{-0.111}$ [−0.198, −0.043]	$\mathbf{0.070}$ [0.057, 0.082]	0.083 [−0.044, 0.209]
Physical activity – θ_t^E	$\mathbf{-0.056}$ [−0.079, −0.033]	0.013 [−0.405, 0.530]	0.006 [−0.037, 0.048]	−0.315 [−1.321, 0.531]

Panel B: Linear production function – Overweight

	Overweight – θ_t^P		**Mental health – θ_t^M**	
	Exog	Endog	Exog	Endog
Overweight (11) – θ_{t-1}^P	$\mathbf{0.639}$ [0.617, 0.661]	$\mathbf{0.626}$ [0.571, 0.679]	$\mathbf{0.033}$ [0.001, 0.064]	0.012 [−0.063, 0.081]
Mental health (11) – θ_{t-1}^M	$\mathbf{-0.009}$ [−0.017, −0.001]	−0.016 [−0.041, 0.009]	$\mathbf{0.925}$ [0.913, 0.936]	$\mathbf{0.930}$ [0.896, 0.967]
Diet quality – θ_t^D	0.004 [−0.004, 0.013]	0.045 [−0.045, 0.141]	$\mathbf{0.070}$ [0.058, 0.082]	0.081 [−0.054, 0.210]
Physical activity – θ_t^E	$\mathbf{-0.063}$ [−0.092, −0.033]	−0.230 [−0.898, 0.408]	0.004 [−0.038, 0.046]	−0.263 [−1.274, 0.569]

Notes: The table presents estimates of the parameters of linear production functions for the entire sample of MCS children. Estimates in the *Exog* columns are obtained without any correction for the potential endogeneity of diet quality and physical activity. Estimates in the *Endog* columns are control function estimates that correct for the endogeneity of diet quality and physical activity. These are obtained using a two-step control function approach: residuals from linear first stage estimates ($\hat{v}^D$, $\hat{v}^E$) are plugged into the second stage linear equation. Panel A uses body fat percentage as a measure of physical health. Panel B uses a binary variable for overweight, derived using age- and sex-specific tables. In both cases, bootstrap 95% confidence intervals with 1000 repetitions are in brackets. *N*: 4,802 (all). Additional controls (not shown) for child non-white ethnicity, socioeconomic status at birth (mother's education, mother's age, family income below 60% of national median, single parent household), health at birth (birthweight), rural household (at age 11), season of interview and number of siblings in household (at age 14).

Chapter 5

Bibliography

Aars, N. A., B. K. Jacobsen, A.-S. Furberg, and S. Grimsgaard (2018). Self-reported physical activity during leisure time was favourably associated with body composition in Norwegian adolescents. *Acta Paediatrica 108*(6), 1122–1127.

Agostinelli, F. and M. Wiswall (2016). Estimating the Technology of Children's Skill Formation. Technical Report w22442, National Bureau of Economic Research, Cambridge, MA.

Ahn, J. V., F. Sera, S. Cummins, and E. Flouri (2018, February). Associations between objectively measured physical activity and later mental health outcomes in children: Findings from the UK Millennium Cohort Study. *Journal of Epidemiology and Community Health 72*(2), 94–100.

Akombi, B. J., K. E. Agho, D. Merom, A. M. Renzaho, and J. J. Hall (2017). Child malnutrition in sub-Saharan Africa: A meta-analysis of demographic and health surveys (2006-2016). *PloS one 12*(5), e0177338.

Alderman, H., J. R. Behrman, S. Grantham-McGregor, F. Lopez-Boo, and S. Urzua (2014, January). Economic perspectives on integrating early child stimulation with nutritional interventions: Economics of ECD interventions. *Annals of the New York Academy of Sciences 1308*(1), 129–138.

Almås, I., O. Attanasio, J. Jalan, F. Oteiza, and M. Vigneri (2018, October). Using data differently and using different data. *Journal of Development Effectiveness 10*(4), 462–481.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality Psychology and Economics. In *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Elsevier.

Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011). Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes. *Review of Economics and Statistics 93*(2), 387–403.

Althabe, F., C. P. Howson, M. Kinney, and J. Lawn (2012). *Born Too Soon: The Global Action Report on Preterm Birth*. World Health Organization. OCLC: 797003676.

Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016). Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, and Program and Social Security Data. *American Economic Journal: Economic Policy 8*(2), 1–43.

Anand, P., J. R. Behrman, H.-A. H. Dang, and S. Jones (2018, October). Varied patterns of catch-up in child growth: Evidence from Young Lives. *Social Science & Medicine 214*, 206–213.

Anderson, M. L. (2008). 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.

Angelucci, M. and V. D. Maro (2015). Program Evaluation and Spillover Effects. Technical report, IZA.

Armand, A., O. Attanasio, P. M. Carneiro, and V. Lechene (2016). The effect of gender-targeted conditional cash transfers on household expenditures: Evidence from a randomized experiment. Technical report, IZA.

Attanasio, O., E. Battistin, and A. Mesnard (2012). Food and Cash Transfers: Evidence from Colombia. *The Economic Journal 122*(559), 92–124.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2018). Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia. Technical Report 1987R2, Cowles Foundation.

Attanasio, O., L. C. Gómez, P. Heredia, and M. Vera-Hernandez (2005). The short-term impact of a conditional cash subsidy on child health and nutrition in Colombia. Technical report, Institute for Fiscal Studies.

Attanasio, O. and V. Lechene (2010). Conditional cash transfers, women and the demand for food. Technical report, Institute for Fiscal Studies.

Attanasio, O., C. Meghir, and E. Nix (2017). Human Capital Development and Parental Investment in India. Technical Report 2026R, Cowles Foundation.

Attanasio, O., C. Meghir, E. Nix, and F. Salvati (2017). Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of Economic Dynamics 25*, 234–259.

Attanasio, O. P. (2015). The determinants of human capital formation during the early years of life: Theory, measurement, and policies. *Journal of the European Economic Association 13*(6), 949–997.

Attanasio, O. P. and V. Lechene (2014). Efficient Responses to Targeted Cash Transfers. *Journal of Political Economy 122*(1), 178–222.

Attanasio, O. P., V. Oppedisano, and M. Vera-Hernández (2015). Should Cash Transfers Be Conditional? Conditionality, Preventive Care, and Health Outcomes. *American Economic Journal: Applied Economics 7*(2), 35–52.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2016, May). Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. Technical Report w22267, National Bureau of Economic Research, Cambridge, MA.

Baird, S., J. A. Bohren, C. McIntosh, and B. Ozler (2015). Designing Experiments to Measure Spillover Effects. Technical Report 15-021, University of Pennsylvania.

Barber, S. L. and P. J. Gertler (2008). The impact of Mexico's conditional cash transfer programme, Oportunidades, on birthweight. *Tropical Medicine & International Health 13*(11), 1405–1414.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, T. Schmidt, and L. Pellerano (2016). Cash transfers: What does the evidence say. Technical report, Overseas Development Institute, London, UK.

Beatty, T. K., L. Blow, T. F. Crossley, and C. O'Dea (2014). Cash by any other name? Evidence on labeling from the UK Winter Fuel Payment. *Journal of Public Economics 118*, 86–96.

Beauducel, A. and P. Y. Herzberg (2006, April). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal 13*(2), 186–203.

Behar, L. and S. Stringfield (1974). A behavior rating scale for the preschool child. *Developmental Psychology 10*(5), 601–610.

Bélanger, M., K. Gray-Donald, J. O'loughlin, G. Paradis, and J. Hanley (2009, March). Influence of Weather Conditions and Season on Physical Activity in Adolescents. *Annals of Epidemiology 19*(3), 180–186.

Belfield, C., A. B. Bowden, A. Klapp, H. Levin, R. Shand, and S. Zander (2015). The Economic Value of Social and Emotional Learning. *Journal of Benefit-Cost Analysis 6*(03), 508–544.

Belloni, A., D.-f. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica 80*(6), 2369–2429.

Belloni, A. and V. Chernozhukov (2013, May). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen (2015). Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education. *American Economic Journal: Economic Policy 7*(3), 86–125.

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin 107*(2), 238–46.

Berglund, L. (1999, December). Latent Variable Analysis of the Rutter Children's Behaviour Questionnaire. *Scandinavian Journal of Educational Research 43*(4), 433–442.

Bhutta, Z. A., J. K. Das, A. Rizvi, M. F. Gaffey, N. Walker, S. Horton, P. Webb, A. Lartey, and R. E. Black (2013). Evidence-based interventions for improvement of maternal and child nutrition: What can be done and at what cost? *The lancet 382*(9890), 452–477.

Biddle, S. J., S. Ciaccioni, G. Thomas, and I. Vergeer (2018, September). Physical activity and mental health in children and adolescents: An updated review of reviews and an analysis of causality. *Psychology of Sport and Exercise 52*, 146–155.

Biddle, S. J. H. and M. Asare (2011, September). Physical activity and mental health in children and adolescents: A review of reviews. *British Journal of Sports Medicine 45*(11), 886–895.

Bitler, M. P. and J. Currie (2005). Does WIC work? The effects of WIC on pregnancy and birth outcomes. *Journal of Policy Analysis and Management 24*(1), 73–91.

Björkman Nyqvist, M. and S. Jayachandran (2017). Mothers Care More, But Fathers Decide: Educating Parents about Child Health in Uganda. *American Economic Review 107*(5), 496–500.

Black, M. M. and K. G. Dewey (2014). Promoting equity through integrated early child development and nutrition interventions: Equity through integrated interventions. *Annals of the New York Academy of Sciences 1308*(1), 1–10.

Blakemore, S.-J. and K. L. Mills (2014, January). Is Adolescence a Sensitive Period for Sociocultural Processing? *Annual Review of Psychology 65*(1), 187–207.

Blanden, J., P. Gregg, and L. Macmillan (2007). Accounting for intergenerational income persistence: Noncognitive skills, ability and education. *The Economic Journal 117*(519), C43–C60.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007, March). Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds. *Econometrica 75*(2), 323–363.

Bor, W., A. J. Dean, J. Najman, and R. Hayatbakhsh (2014, July). Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Australian & New Zealand Journal of Psychiatry 48*(7), 606–616.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources 43*(4), 972–1059.

Bowles, S., H. Gintis, and M. Osborne (2001, December). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature 39*(4), 1137–1176.

Briend, A., T. Khara, and C. Dolan (2015). Wasting and Stunting—Similarities and Differences: Policy and Programmatic Implications. *Food and Nutrition Bulletin 36*(1_suppl1), S15–S23.

Butler, N. B. (2016a). 1970 British Cohort Study: Five-Year Follow-Up, 1975.

Butler, N. B. (2016b). 1970 British Cohort Study: Ten-Year Follow-Up, 1980.

Butler, N. B. (2017). 1970 British Cohort Study: Sixteen-Year Follow-Up, 1986.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014). Early Childhood Investments Substantially Boost Adult Health. *Science 343*(6178), 1478–1485.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry 36*(1), 113–149.

Carneiro, P., C. Meghir, and M. Parey (2013, January). Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association 11*, 123–160.

Case, A., A. Fertig, and C. Paxson (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics 24*(2), 365–389.

Cawley, J. (2015, September). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics 43*, 244–268.

Central Bank of Nigeria (2017). Data & Statistics. https://www.cbn.gov.ng/rates/.

Chamberlain, R. C. (2013). 1970 British Cohort Study: Birth and 22-Month Subsample, 1970-1972.

Chen, S., M. Schreiner, and G. Woller (2008). Simple Poverty Scorecard Poverty Assessment Tool Nigeria.

Chernozhukov, V., C. Hansen, and M. Spindler (2016). High-dimensional metrics in R.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011, November). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Cheung, G. W. and R. S. Lau (2012, April). A Direct Comparison Approach for Testing Measurement Invariance. *Organizational Research Methods 15*(2), 167–198.

Cheung, G. W. and R. B. Rensvold (2002, April). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal 9*(2), 233–255.

Chiteji, N. (2010, May). Time Preference, Noncognitive Skills and Well Being across the Life Course: Do Noncognitive Skills Encourage Healthy Behavior? *American Economic Review 100*(2), 200–204.

Cobb-Clark, D. A., S. C. Kassenboehmer, and S. Schurer (2014, February). Healthy habits: The connection between diet, exercise, and locus of control. *Journal of Economic Behavior & Organization 98*, 1–28.

Cole, T. J. and T. Lobstein (2012, August). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity: Extended international BMI cut-offs. *Pediatric Obesity 7*(4), 284–294.

Collishaw, S. (2015, March). Annual Research Review: Secular trends in child and adolescent mental health. *Journal of Child Psychology and Psychiatry 56*(3), 370–393.

Collishaw, S., B. Maughan, R. Goodman, and A. Pickles (2004, November). Time trends in adolescent mental health. *Journal of Child Psychology and Psychiatry 45*(8), 1350–1362.

Collishaw, S., B. Maughan, L. Natarajan, and A. Pickles (2010, March). Trends in adolescent emotional problems in England: A comparison of two national cohorts twenty years apart: Twenty-year trends in emotional problems. *Journal of Child Psychology and Psychiatry 51*(8), 885–894.

Colquitt, J. L., E. Loveman, C. O'Malley, L. B. Azevedo, E. Mead, L. Al-Khudairy, L. J. Ells, M.-I. Metzendorf, and K. Rees (2016, March). Diet, physical activity, and behavioural interventions for the treatment of overweight or obesity in preschool children up to the age of 6 years. *Cochrane Database of Systematic Reviews* (3).

Connelly, R. and L. Platt (2014, December). Cohort Profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology 43*(6), 1719–1725.

Conti, G., S. Frühwirth-Schnatter, J. J. Heckman, and R. Piatek (2014, November). Bayesian exploratory factor analysis. *Journal of Econometrics 183*(1), 31–57.

Conti, G. and C. Hansman (2013, March). Personality and the education–health gradient: A note on "Understanding differences in health behaviors by education". *Journal of Health Economics 32*(2), 480–485.

Conti, G., M. Hanson, H. Inskip, S. Crozier, C. Cooper, and K. Godfrey (2018). Beyond Birth Weight: The Origins of Human Capital. Technical Report W18/30, Institute for Fiscal Studies.

Conti, G., J. Heckman, and S. Urzua (2010, May). The Education-Health Gradient. *American Economic Review 100*(2), 234–238.

Conti, G. and J. J. Heckman (2010, September). Understanding the Early Origins of the Education–Health Gradient: A Framework That Can Also Be Applied to Analyze Gene–Environment Interactions. *Perspectives on Psychological Science 5*(5), 585–605.

Conti, G., J. J. Heckman, and R. Pinto (2016). The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour. *The Economic Journal 126*(596), F28–F65.

Conti, G., J. J. Heckman, and S. Urzua (2011). Early endowments, education, and health. Technical Report 2011-01, Human Capital and Economic Opportunity Global Working Group.

Corman, H., D. M. Dave, and N. E. Reichman (2018). Effect of prenatal care on birth outcomes: Reconciling a messy literature. Technical Report 24885, National Bureau of Economic Research.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education*, Volume 1, pp. 697–812. Elsevier.

Cunha, F., J. J. Heckman, and S. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Currie, J., M. Neidell, and J. F. Schmieder (2009). Air pollution and infant health: Lessons from New Jersey. *Journal of Health Economics 28*(3), 688–703.

Daidone, S., B. Davis, S. Handa, and P. Winters (2017). The household- and individual-level economic impacts of cash transfer programmes in Sub-Saharan Africa. Technical report, Food and Agriculture Organisation.

Debes, A. K., A. Kohli, N. Walker, K. Edmond, and L. C. Mullany (2013). Time to initiation of breastfeeding and neonatal mortality and morbidity: A systematic review. *BMC Public Health 13*(Suppl 3), S19.

Del Boca, D., C. Flinn, and M. Wiswall (2014, January). Household Choices and Child Development. *The Review of Economic Studies 81*(1), 137–185.

Del Boca, D., C. Monfardini, and C. Nicoletti (2017). Parental and child time investments and the cognitive development of adolescents. *Journal of Labor Economics 35*(2), 565–608.

Dell, M., B. F. Jones, and B. A. Olken (2014, September). What Do We Learn from the Weather? The New Climate-Economy Literature †. *Journal of Economic Literature 52*(3), 740–798.

Deming, D. J. (2017, November). The Growing Importance of Social Skills in the Labor Market*. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Devaney, B., L. Bilheimer, and J. Schore (1992). Medicaid Costs and Birth Outcomes: The Effects of Prenatal WIC Participation and the Use of Prenatal Care. *Journal of Policy Analysis and Management 11*(4), 573.

Dewey, K. G. and S. Adu-Afarwuah (2008). Systematic review of the efficacy and effectiveness of complementary feeding interventions in developing countries. *Maternal & Child Nutrition 4*(s1), 24–85.

Dewey, K. G. and K. Begum (2011). Long-term consequences of stunting in early life. *Maternal & Child Nutrition 7*, 5–18.

Dickey, W. C. and S. J. Blumberg (2004, September). Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry 43*(9), 1159–1167.

Doyle, O. (2017). The First 2,000 Days and Child Skills: Evidence from a Randomized Experiment of Home Visiting. Technical report, Geary Institute, University College Dublin.

Duncan, J. S., W. G. Hopkins, G. Schofield, and E. K. Duncan (2008, August). Effects of Weather on Pedometer-Determined Physical Activity in Children:. *Medicine & Science in Sports & Exercise 40*(8), 1432–1438.

Durlak, J. A., R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger (2011, January). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions: Social and Emotional Learning. *Child Development 82*(1), 405–432.

Durlak, J. A., R. P. Weissberg, and M. Pachan (2010, June). A Meta-Analysis of After-School Programs That Seek to Promote Personal and Social Skills in Children and Adolescents. *American Journal of Community Psychology 45*(3-4), 294–309.

Elshibly, E. M. and G. Schmalisch (2008, December). The effect of maternal anthropometric characteristics and social factors on gestational age and birth weight in Sudanese newborn infants. *BMC Public Health 8*(244), 1–7.

Elzhov, T., K. M. Mullen, A.-N. Spiess, and B. Bolker (2016). R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds.

Evans, W. N. and D. S. Lien (2005). The benefits of prenatal care: Evidence from the PAT bus strike. *Journal of Econometrics 125*(1-2), 207–239.

Federal Ministry of Health (2014). National Strategic Plan of Action for Nutrition (2014-2019).

Fernald, L. C., P. J. Gertler, and L. M. Neufeld (2008). Role of cash in conditional cash transfer programmes for child health, growth, and development: An analysis of Mexico's Oportunidades. *The Lancet 371*(9615), 828–837.

Fernald, L. C. H., P. Kariger, M. Hidrobo, and P. J. Gertler (2012). Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences 109*(Supplement_2), 17273–17280.

Fernald, L. C. H., E. Prado, P. Kariger, and A. Raikes (2017). A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries. Technical report, World Bank.

FEWS (2016). Nigeria Food Security Outlook Update - September 2016. Technical report, Famine Early Warning Systems Network.

Fink, E., P. Patalay, H. Sharpe, S. Holley, J. Deighton, and M. Wolpert (2015, May). Mental Health Difficulties in Early Adolescence: A Comparison of Two Cross-Sectional Studies in England From 2009 to 2014. *Journal of Adolescent Health 56*(5), 502–507.

Fitzsimons, E., B. Malde, A. Mesnard, and M. Vera-Hernández (2016). Nutrition, information and household behavior: Experimental evidence from Malawi. *Journal of Development Economics 122*, 113–126.

Flora, D. B. and P. J. Curran (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods 9*(4), 466–491.

Fowler, P. C. and R. M. Park (1979, October). Factor Structure of the Preschool Behavior Questionnaire in a Normal Population. *Psychological Reports 45*(2), 599–606.

Fox, J. (2016). Package 'polycor'.

Frongillo, E. A., P. H. Nguyen, K. K. Saha, T. Sanghvi, K. Afsana, R. Haque, J. Baker, M. T. Ruel, R. Rawat, and P. Menon (2017). Large-Scale Behavior-Change Initiative for Infant and Young Child Feeding Advanced Language and Motor Development in a Cluster-Randomized Program Evaluation in Bangladesh. *The Journal of Nutrition 147*(2), 256–263.

Fuhrmann, D., L. J. Knoll, and S.-J. Blakemore (2015, October). Adolescence as a Sensitive Period of Brain Development. *Trends in Cognitive Sciences 19*(10), 558–566.

Gaarder, M. M., A. Glassman, and J. E. Todd (2010). Conditional cash transfers and health: Unpacking the causal chain. *Journal of Development Effectiveness 2*(1), 6–50.

Galasso, E. and A. Wagstaff (2018). The Aggregate Income Losses from Childhood Stunting and the Returns to a Nutrition Intervention Aimed at Reducing Stunting. Technical report, World Bank.

Gaysina, D., D. M. Fergusson, L. D. Leve, J. Horwood, D. Reiss, D. S. Shaw, K. K. Elam, M. N. Natsuaki, J. M. Neiderhiser, and G. T. Harold (2013, September). Maternal Smoking During Pregnancy and Offspring Conduct Problems: Evidence From 3 Independent Genetically Sensitive Research Designs. *JAMA Psychiatry 70*(9), 956.

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics 34*(2), 509–543.

Gertler, P. J., S. W. Martinez, and M. Rubio-Codina (2012). Investing Cash Transfers to Raise Long-Term Living Standards. *American Economic Journal: Applied Economics 4*(1), 164–192.

Ghosh, S. (2016). Protein Quality in the First Thousand Days of Life. *Food and Nutrition Bulletin 37*(1_suppl), S14–S21.

Gilbert, E., A. Conolly, L. Calderwood, and N. Rose (2017). Measuring young people's physical activity using accelerometers in the UK Millennium Cohort Study. Technical Report 15, Centre for Longitudinal Studies, Institute of Education.

Goodman, A. and R. Goodman (2011, January). Population mean scores predict child mental disorder rates: Validating SDQ prevalence estimators in Britain: SDQ prevalence estimators. *Journal of Child Psychology and Psychiatry 52*(1), 100–108.

Goodman, A., H. E. Joshi, B. Nasim, and C. Tyler (2015). Social and emotional skills in childhood and their long-term effects on adult life. Technical report, Institute of Education.

Goodman, A., D. L. Lamping, and G. B. Ploubidis (2010, November). When to Use Broader Internalising and Externalising Subscales Instead of the Hypothesised Five Subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British Parents, Teachers and Children. *Journal of Abnormal Child Psychology 38*(8), 1179–1191.

Goodman, R. (1994). A modified version of the Rutter parent questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry 35*(8), 1483–1494.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of child psychology and psychiatry 38*(5), 581–586.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp, and I. C. D. S. Group (2007). Developmental potential in the first 5 years for children in developing countries. *The lancet 369*(9555), 60–70.

Grantham-McGregor, S., C. Powell, S. Walker, and J. Himes (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican Study. *The Lancet 338*(8758), 1–5.

Grantham-McGregor, S. M., L. C. H. Fernald, R. M. C. Kagawa, and S. Walker (2014). Effects of integrated child development and nutrition interventions on child development and nutritional status: Integrated child development interventions. *Annals of the New York Academy of Sciences 1308*(1), 11–32.

Griffith, R., M. O'Connell, and K. Smith (2015, March). Relative prices, consumer preferences, and the demand for food. *Oxford Review of Economic Policy 31*(1), 116–130.

Griffiths, L., M. Geraci, M. Cortina-Borja, F. Sera, C. Law, H. Joshi, A. Ness, and C. Dezateux (2016, April). Associations between children's behavioural and emotional development and objectively measured physical activity and sedentary time: Findings from the UK Millennium Cohort Study. *Longitudinal and Life Course Studies 7*(2), 124–143.

Hamer, M. and E. Stamatakis (2018, June). Relative proportion of vigorous physical activity, total volume of moderate to vigorous activity, and body mass index in youth: The Millennium Cohort Study. *International Journal of Obesity 42*(6), 1239–1242.

Handa, S., L. Natali, D. Seidenfeld, G. Tembo, and B. Davis (2018). Can unconditional cash transfers raise long-term living standards? Evidence from Zambia. *Journal of Development Economics 133*, 42–65.

Handa, S., D. Seidenfeld, B. Davis, G. Tembo, and Zambia Cash Transfer Evaluation Team (2016). The Social and Productive Impacts of Zambia's Child Grant: Impacts of Zambia's Child Grant. *Journal of Policy Analysis and Management 35*(2), 357–387.

Hansford, F., A. Visram, E. Jones, and P. Ward (2017). Integrated Evaluation Report of the WINNN Programme. Technical report, Oxford Policy Management, Oxford, UK.

Harding, M. and M. Lovenheim (2017, May). The effect of prices on nutrition: Comparing the impact of product- and nutrient-specific taxes. *Journal of Health Economics 53*, 53–71.

Haroon, S., J. K. Das, R. A. Salam, A. Imdad, and Z. A. Bhutta (2013). Breastfeeding promotion interventions and breastfeeding practices: A systematic review. *BMC Public Health 13*(Suppl 3), S20.

Harris, K. C., L. K. Kuramoto, M. Schulzer, and J. E. Retallack (2009, March). Effect of school-based physical activity interventions on body mass index in children: A meta-analysis. *Canadian Medical Association Journal 180*(7), 719–726.

Harrison, F., A. Goodman, E. M. F. van Sluijs, L. B. Andersen, G. Cardon, R. Davey, K. F. Janz, S. Kriemler, L. Molloy, A. S. Page, R. Pate, J. J. Puder, L. B. Sardinha, A. Timperio, N. Wedderkopp, and A. P. Jones (2017, December). Weather and children's physical activity; how and why do relationships vary between countries? *International Journal of Behavioral Nutrition and Physical Activity 14*(1), 14–74.

Harrison, F., A. P. Jones, G. Bentham, E. M. van Sluijs, A. Cassidy, and S. J. Griffin (2011). The impact of rainfall and school break time policies on physical activity in 9-10 year old British children: A repeated measures study. *International Journal of Behavioral Nutrition and Physical Activity 8*(1), 47.

Harrison, F., E. M. van Sluijs, K. Corder, U. Ekelund, and A. Jones (2015). The changing relationship between rainfall and children's physical activity in spring and summer: A longitudinal study. *International Journal of Behavioral Nutrition and Physical Activity 12*(1), 12–41.

Headey, D. and G. Palloni (2018). Water, Sanitation and Child Health. Technical Report 01753, International Food Policy Research Institute.

Headey, D. D., K. Hirvonen, and J. F. Hoddinott (2017). Animal sourced foods and child stunting. Technical Report 01695, International Food Policy Research Institute.

Heckman, J., J. E. Humphries, and G. Veramendi (2016, May). Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking. Technical Report w22291, National Bureau of Economic Research, Cambridge, MA.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Hirvonen, K. (2014, January). Measuring catch-up growth in malnourished populations. *Annals of Human Biology 41*(1), 67–75.

Hoddinott, J., J. A. Maluccio, J. R. Behrman, R. Flores, and R. Martorell (2008). Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *The lancet 371*(9610), 411–416.

Hollis, D. and M. McCarty (2017). UKCP09: Met Office gridded and regional land surface climate observation datasets.

Horn, J. L. (1965, June). A rationale and test for the number of factors in factor analysis. *Psychometrika 30*(2), 179–185.

Hoynes, H., D. Miller, and D. Simon (2015). Income, the Earned Income Tax Credit, and Infant Health. *American Economic Journal: Economic Policy 7*(1), 172–211.

Hoynes, H., M. Page, and A. H. Stevens (2011). Can targeted transfers improve birth outcomes? *Journal of Public Economics 95*(7-8), 813–827.

Hsieh, C.-S. and H. van Kippersluis (2018, July). Smoking initiation: Peers and personality. *Quantitative Economics 9*(2), 825–863.

Imdad, A., M. Y. Yakoob, and Z. A. Bhutta (2011). Effect of breastfeeding promotion interventions on breastfeeding rates, with special focus on developing countries. *BMC Public Health 11*(Suppl 3), S24.

Jacob, B., L. Lefgren, and E. Moretti (2007). The Dynamics of Criminal Behavior: Evidence from Weather Shocks. *Journal of Human Resources XLII*(3), 489–527.

Janssen, I., P. T. Katzmarzyk, W. F. Boyce, C. Vereecken, C. Mulvihill, C. Roberts, C. Currie, W. Pickett, and The Health Behaviour in School-Aged Children Obesity Working Group* (2005, May). Comparison of overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns. *Obesity Reviews 6*(2), 123–132.

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics 7*(1), 63–88.

Jennings, A., A. Welch, E. M. F. van Sluijs, S. J. Griffin, and A. Cassidy (2011, March). Diet Quality Is Independently Associated with Weight Status in Children Aged 9–10 Years. *The Journal of Nutrition 141*(3), 453–459.

Jones, N. R. V., A. I. Conklin, M. Suhrcke, and P. Monsivais (2014, October). The Growing Price Gap between More and Less Healthy Foods: Analysis of a Novel Longitudinal UK Dataset. *PLoS ONE 9*(10), e109343.

Joyce, T., D. Gibson, and S. Colman (2005). The changing association between prenatal participation in WIC and birth outcomes in New York City. *Journal of Policy Analysis and Management 24*(4), 661–685.

Joyce, T., A. Racine, and C. Yunzal-Butler (2008). Reassessing the WIC effect: Evidence from the Pregnancy Nutrition Surveillance System. *Journal of Policy Analysis and Management 27*(2), 277–303.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement 20*(1), 141–151.

Kautz, T., J. J. Heckman, R. Diris, B. Ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, National Bureau of Economic Research.

Kim, S. S., P. H. Nguyen, L. M. Tran, T. Sanghvi, Z. Mahmud, M. R. Haque, K. Afsana, E. A. Frongillo, M. T. Ruel, and P. Menon (2018). Large-Scale Social and Behavior Change Communication Interventions Have Sustained Impacts on Infant and Young Child Feeding Knowledge and Practices: Results of a 2-Year Follow-Up Study in Bangladesh. *The Journal of Nutrition 148*, 1–10.

Kim, S. S., R. Rawat, E. M. Mwangi, R. Tesfaye, Y. Abebe, J. Baker, E. A. Frongillo, M. T. Ruel, and P. Menon (2016). Exposure to Large-Scale Social and Behavior Change Communication Interventions Is Associated with Improvements in Infant and Young Child Feeding Practices in Ethiopia. *PLOS ONE 11*(10), e0164800.

Klein, J. M., A. Gonçalves, and C. F. Silva (2009). The Rutter Children Behaviour Questionnaire for teachers: From psychometrics to norms, estimating caseness. *Psico-USF 14*(2), 157–165.

Kohlboeck, G., S. Sausenthaler, M. Standl, S. Koletzko, C.-P. Bauer, A. von Berg, D. Berdel, U. Krämer, B. Schaaf, I. Lehmann, O. Herbarth, J. Heinrich, and for the GINI-plus and LISA-plus Study Groups (2012). Food Intake, Diet Quality and Behavioral Problems in Children: Results from the GINI-plus/LISA-plus Studies. *Annals of Nutrition and Metabolism 60*(4), 247–256.

Kramer, M. S. and R. Kakuma (2012). Optimal duration of exclusive breastfeeding. *Cochrane Database of Systematic Reviews* (8).

Laidley, T. and D. Conley (2018, September). The Effects of Active and Passive Leisure on Cognition in Children: Evidence from Exogenous Variation in Weather. *Social Forces 97*(1), 129–156.

Lassi, Z. S., J. K. Das, G. Zahid, A. Imdad, and Z. A. Bhutta (2013). Impact of education and provision of complementary feeding on growth and morbidity in children less than 2 years of age in developing countries: A systematic review. *BMC Public Health 13*(Suppl 3), S13.

Laverty, A. A., L. Magee, C. A. Monteiro, S. Saxena, and C. Millett (2015). Sugar and artificially sweetened beverage consumption and adiposity changes: National longitudinal study. *International Journal of Behavioral Nutrition and Physical Activity 12*(137), 1–10.

Layard, R., A. E. Clark, F. Cornaglia, N. Powdthavee, and J. Vernoit (2014, November). What Predicts a Successful Life? A Life-course Model of Well-being. *The Economic Journal 124*(580), F720–F738.

Leroy, J. L. (2011). Zscore06: Stata command for the calculation of anthropometric z-scores using the 2006 WHO child growth standards.

Leroy, J. L., M. Ruel, and E. Verhofstadt (2009). The impact of conditional cash transfer programmes on child nutrition: A review of evidence using a programme theory framework. *Journal of Development Effectiveness 1*(2), 103–129.

Levere, M., G. Acharya, and P. Bharadwaj (2016). The role of information and cash transfers on early childhood development: Evidence from Nepal. Technical report, World Bank.

Li, C.-H. (2016, September). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods 48*(3), 936–949.

Li, K. and D. J. Poirier (2003). Bayesian analysis of an econometric model of birth inputs and outputs. *Journal of Population Economics 16*(3), 597–625.

Lindqvist, E. and R. Vestman (2011, January). The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics 3*(1), 101–128.

Linnemayr, S. and H. Alderman (2011). Almost random: Evaluating a large-scale randomized nutrition program in the presence of crossover. *Journal of Development Economics 96*(1), 106–114.

Macours, K., N. Schady, and R. Vakis (2012). Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment. *American Economic Journal: Applied Economics 4*(2), 247–273.

Maluccio, J. A., J. Hoddinott, J. R. Behrman, R. Martorell, A. R. Quisumbing, and A. D. Stein (2009). The impact of improving nutrition during early childhood on education among Guatemalan adults. *The Economic Journal 119*(537), 734–763.

Manley, J., S. Gitter, and V. Slavchevska (2013). How Effective are Cash Transfers at Improving Nutritional Status? *World Development 48*, 133–155.

Maynard, L. M., W. Wisemandle, A. F. Roche, W. C. Chumlea, S. S. Guo, and R. M. Siervogel (2001, February). Childhood Body Composition in Relation to Body Mass Index. *Pediatrics 107*(2), 344–350.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of classification 6*(1), 97–103.

McGee, R., S. Williams, J. Bradshaw, J. L. Chapel, A. Robins, and P. A. Silva (1985). The Rutter scale for completion by teachers: Factor structure and relationships with cognitive abilities and family adversity for a sample of New Zealand children. *Journal of Child Psychology and Psychiatry 26*(5), 727–739.

McGovern, M. E., A. Krishna, V. M. Aguayo, and S. Subramanian (2017). A review of the evidence linking child stunting to economic outcomes. *International Journal of Epidemiology 46*(4), 1171–1191.

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

Mead, E., T. Brown, K. Rees, L. B. Azevedo, V. Whittaker, D. Jones, J. Olajide, G. M. Mainardi, E. Corpeleijn, C. O'Malley, E. Beardsmore, L. Al-Khudairy, L. Baur, M.-I. Metzendorf, A. Demaio, and L. J. Ells (2017). Diet, physical activity and behavioural interventions for the treatment of overweight or obese children from the age of 6 to 11 years. *Cochrane Database of Systematic Reviews* (6).

Meade, A. W., E. C. Johnson, and P. W. Braddy (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology 93*(3), 568–592.

Mendolia, S. and I. Walker (2014, September). The effect of noncognitive traits on health behaviours in adolescence. *Health Economics 23*(9), 1146–1158.

Menon, P., P. H. Nguyen, K. K. Saha, A. Khaled, A. Kennedy, L. M. Tran, T. Sanghvi, N. Hajeebhoy, J. Baker, S. Alayon, K. Afsana, R. Haque, E. A. Frongillo, M. T. Ruel, and R. Rawat (2016). Impacts on Breastfeeding Practices of At-Scale Strategies That Combine Intensive Interpersonal Counseling, Mass Media, and Community Mobilization: Results of Cluster-Randomized Program Evaluations in Bangladesh and Viet Nam. *PLOS Medicine 13*(10), e1002159.

Millsap, R. E. and J. Yun-Tein (2004, July). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research 39*(3), 479–515.

Munro, A., B. Kebede, M. Tarazona-Gomez, and A. Verschoor (2011). The lion's share: An experimental analysis of polygamy in northern Nigeria. Technical Report 10-27, National Graduate Institute for Policy Studies.

Muthén, B. (1984, March). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika 49*(1), 115–132.

Muthen, B. O., S. H. C. du Toit, and D. Spisic (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Technical report, Unpublished manuscipt.

Ness, A. R., S. D. Leary, C. Mattocks, S. N. Blair, J. J. Reilly, J. Wells, S. Ingle, K. Tilling, G. D. Smith, and C. Riddoch (2007, March). Objectively Measured Physical Activity and Fat Mass in a Large Cohort of Children. *PLoS Medicine 4*(3), e97.

Nevill, A. M., A. D. Stewart, T. Olds, and R. Holder (2006, January). Relationship between adiposity and body size reveals limitations of BMI. *American Journal of Physical Anthropology 129*(1), 151–156.

NHS (2017). Mental Health of Children and Young People in England, 2017. Technical report, National Health Service.

NHS (2018). Statistics on Obesity, Physical Activity and Diet - England, 2018. Technical report, National Health Service.

NPC and ICF (2014). Nigeria Demographic and Health Survey 2013. Technical report, National Population Commission and ICF International, Abuja, Nigeria.

Oddy, W. H., M. Robinson, G. L. Ambrosini, T. A. O′Sullivan, N. H. de Klerk, L. J. Beilin, S. R. Silburn, S. R. Zubrick, and F. J. Stanley (2009, July). The association between dietary patterns and mental health in early adolescence. *Preventive Medicine 49*(1), 39–44.

OECD (1982). *The OECD List of Social Indicators*. Paris: Organisation for Economic Cooperation and Development.

Okeke, T. C., J. O. Enwereji, H. U. Ezegwui, A. C. Adiuku-Brown, C. O. Adiri, and O. A. Okezie (2014). Prospective Estimation of Duration of Pregnancy in Women Attending Antenatal Care at the University of Nigeria Teaching Hospital, Enugu, Nigeria. *Annual Research & Review in Biology 4*(14), 2274.

Omigbodun, A. O. and A. Adewuyi (1997). Duration of human singleton pregnancies in Ibadan, Nigeria. *Journal of the National Medical Association 89*, 617–621.

O'Neil, A., S. E. Quirk, S. Housden, S. L. Brennan, L. J. Williams, J. A. Pasco, M. Berk, and F. N. Jacka (2014, October). Relationship Between Diet and Mental Health in Children and Adolescents: A Systematic Review. *American Journal of Public Health 104*(10), e31–e42.

ONS (2018). Relative regional consumer price levels of goods and services, UK: 2016. Technical report, Office for National Statistics.

Pace, N. and L. Pellerano (2016). Does 'Soft Conditionality' Increase the Impact of Cash Transfers on Desired Outcomes? Evidence from a Randomized Control Trial in Lesotho. Technical Report 33, Department of Economics - University of Cà Foscari.

Paxson, C. and N. Schady (2010). Does Money Matter? The Effects of Cash Transfers on Child Development in Rural Ecuador. *Economic Development and Cultural Change 59*(1), 187–229.

Penny, M. E., H. M. Creed-Kanashiro, R. C. Robert, M. R. Narro, L. E. Caulfield, and R. E. Black (2005). Effectiveness of an educational intervention delivered through the health services to improve nutrition in young children: A cluster-randomised controlled trial. *The Lancet 365*(9474), 1863–1872.

Perry, M. and D. Hollis (2005, June). The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology 25*(8), 1041–1054.

Perumal, N., D. G. Bassani, and D. E. Roth (2018). Use and Misuse of Stunting as a Measure of Child Health. *The Journal of Nutrition 148*(3), 311–315.

Plewis, I., L. Calderwood, D. Hawkes, G. Hughes, and H. Joshi (2007). Millennium cohort study: Technical report on sampling. Technical report, Centre for Longitudinal Studies, Institute of Education, London, UK.

Pot, G. K., M. Richards, C. J. Prynne, and A. M. Stephen (2014, December). Development of the Eating Choices Index (ECI): A four-item index to measure healthiness of diet. *Public Health Nutrition 17*(12), 2660–2666.

Prentice, A. M., K. A. Ward, G. R. Goldberg, L. M. Jarjou, S. E. Moore, A. J. Fulford, and A. Prentice (2013). Critical windows for nutritional interventions against stunting. *American Journal of Clinical Nutrition 97*(5), 911–918.

Puentes, E., F. Wang, J. R. Behrman, F. Cunha, J. Hoddinott, J. A. Maluccio, L. S. Adair, J. B. Borja, R. Martorell, and A. D. Stein (2016, September). Early life height and weight production functions with endogenous energy and protein inputs. *Economics & Human Biology 22*, 65–81.

Putnick, D. L. and M. H. Bornstein (2016, September). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review 41*, 71–90.

Raîche, G., T. A. Walls, D. Magis, M. Riopel, and J.-G. Blais (2013, January). Non-Graphical Solutions for Cattell's Scree Test. *Methodology 9*(1), 23–29.

Ramires, V. V., S. C. Dumith, F. C. Wehrmeister, P. C. Hallal, A. M. B. Menezes, and H. Gonçalves (2016, December). Physical activity throughout adolescence and body composition at 18 years: 1993 Pelotas (Brazil) birth cohort study. *International Journal of Behavioral Nutrition and Physical Activity 13*(1), 13:105.

Reardon, S. F. and X. A. Portilla (2016, July). Recent Trends in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry. *AERA Open 2*(3), 233285841665734.

Reichert, F. F., J. C. K. Wells, U. Ekelund, A. M. B. Menezes, C. G. Victora, and P. C. Hallal (2015, June). Prospective Associations between Physical Activity Level and Body Composition in Adolescence: 1993 Pelotas (Brazil) Birth Cohort. *Journal of Physical Activity and Health 12*(6), 834–839.

Revelle, W. (2018). Psych: Procedures for Personality and Psychological Research. Northwestern University.

Revelle, W. and T. Rocklin (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research 14*(4), 403–414.

Rich, C., M. Cortina-Borja, C. Dezateux, M. Geraci, F. Sera, L. Calderwood, H. Joshi, and L. J. Griffiths (2013). Predictors of non-response in a UK-wide cohort study of children's accelerometer-determined physical activity using postal methods. *BMJ Open 3*(3), e002290.

Roantree, B. and K. Vira (2018). The rise and rise of women's employment in the UK. Technical Report BN234, Institute for Fiscal Studies.

Roelen, K., S. Devereux, A.-G. Abdulai, B. Martorano, T. Palermo, and L. P. Ragno (2017). How to Make 'Cash Plus' Work: Linking Cash Transfers to Services and Sectors. Technical report, UNICEF.

Rosenberg, R. E., A. N. U. Ahmed, S. Ahmed, S. K. Saha, M. A. Chowdhury, R. E. Black, M. Santosham, and G. L. Darmstadt (2009). Determining Gestational Age in a Low-resource Setting: Validity of Last Menstrual Period. *Journal of Health, Population and Nutrition 27*(3), 332–338.

Rosenzweig, M. R. and T. P. Schultz (1982). The Behavior of Mothers as Inputs to Child Health: The Determinants of Birth Weight, Gestation, and Rate of Fetal Growth. In V. R. Fuchs (Ed.), *Economic Aspects of Health*. Chicago: University of Chicago Press.

Rosseel, Y. (2012). Lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software 48*(2), 1–36.

Roy, S. K., G. J. Fuchs, Z. Mahmud, G. Ara, S. Islam, S. Shafique, S. S. Akter, and B. Chakraborty (2005). Intensive nutrition education with or without supplementary feeding improves the nutritional status of moderately-malnourished children in Bangladesh. *Journal of Health, Population and Nutrition 23*(4), 320–330.

Roy, S. K., S. P. Jolly, S. Shafique, G. J. Fuchs, Z. Mahmud, B. Chakraborty, and S. Roy (2007). Prevention of malnutrition among young children in rural Bangladesh by a food-health-care educational intervention: A randomized, controlled trial. *Food and nutrition bulletin 28*(4), 375–383.

Rubio-Codina, M., M. C. Araujo, O. Attanasio, P. Muñoz, and S. Grantham-McGregor (2016). Concurrent Validity and Feasibility of Short Tests Currently Used to Measure Early Childhood Development in Large Scale Studies. *PLOS ONE 11*(8), e0160962.

Rutter, M., J. Tizard, and K. Whitmore (1970). *Education, Health, and Behaviour*. London, UK: Prentice Hall.

Sass, D. A., T. A. Schmitt, and H. W. Marsh (2014, April). Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Structural Equation Modeling: A Multidisciplinary Journal 21*(2), 167–180.

Savelyev, P. and K. T. K. Tan (2019). Socioemotional Skills, Education, and Health-Related Outcomes of High-Ability Individuals. *American Journal of Health Economics 5*(2), 250–280.

Schady, N. and J. Rosero (2008). Are cash transfers made to women spent like other sources of income? *Economics Letters 101*(3), 246–248.

Schreiner, M. (2015). Simple Poverty Scorecard® Poverty-Assessment Tool Nigeria.

Segal, C. (2008). Classroom behavior. *Journal of Human Resources 43*(4), 783–814.

Segal, C. (2013, August). Misbehavior, Education, and Labor Market Outcomes. *Journal of the European Economic Association 11*(4), 743–779.

Semba, R. D., M. Shardell, F. A. Sakr Ashour, R. Moaddel, I. Trehan, K. M. Maleta, M. I. Ordiz, K. Kraemer, M. A. Khadeer, L. Ferrucci, and M. J. Manary (2016). Child Stunting is Associated with Low Circulating Essential Amino Acids. *EBioMedicine 6*, 246–252.

Sera, F., M. Cortina-Borja, L. Griffiths, and C. Dezateux (2013, November). Effects of physical activity on reduction of adiposity across the entire obesity distribution in primary school-aged children: Findings from the Millennium Cohort Study. *The Lancet 382*, S89.

Sharp, K. and A. Cornelius (2015). Child Development Grant Programme Evaluation Qualitative Midline Report. Technical report, ePact Consortium.

Sharp, K., A. Cornelius, and V. Gadhavi (2018). Child Development Grant Programme Evaluation Qualitative Endline Report. Technical report, ePact Consortium.

Sharp, K., A. Visram, G. Bahety, and A. Kardan (2016). Child Development Grant Programme Evaluation Final Process Evaluation Report Round I. Technical report, ePact Consortium.

Shrimpton, R., C. G. Victora, M. de Onis, R. C. Lima, M. Blössner, and G. Clugston (2001). Worldwide timing of growth faltering: Implications for nutritional interventions. *Pediatrics 107*(5), e75–e75.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton: Chapman & Hall/CRC.

Solivetti, L. M. (1994). Family, marriage and divorce in a Hausa community: A sociological model. *Africa 64*(02), 252–271.

Squires, J. and D. D. Bricker (2009). *Ages & Stages Questionnaires*. Brookes Publishing Company. OCLC: 1010500088.

Steiger, J. H. (1989). EzPATH Causal modeling.

Steinberg, L. D. (2015). *Age of Opportunity: Lessons from the New Science of Adolescence* (First Mariner books edition ed.). An Eamon Dolan Book. Boston: Houghton Mifflin Harcourt. OCLC: 1048394932.

Stock, J. H. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge University Press.

Stone, L. L., R. Otten, R. C. M. E. Engels, A. A. Vermulst, and J. M. A. M. Janssens (2010, September). Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review. *Clinical Child and Family Psychology Review 13*(3), 254–274.

Tremblay, R. E., L. Desmarais-Gervais, C. Gagnon, and P. Charlebois (1987, December). The Preschool Behaviour Questionnaire: Stability of its Factor Structure Between Cultures, Sexes, Ages and Socioeconomic Classes. *International Journal of Behavioral Development 10*(4), 467–484.

Tucker, P. and J. Gilliland (2007, December). The effect of season and weather on physical activity: A systematic review. *Public Health 121*(12), 909–922.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2016c). 1970 British Cohort Study: Forty-Two-Year Follow-Up, 2012.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2016b). 1970 British Cohort Study: Thirty-Eight-Year Follow-Up, 2008-2009.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2016a). 1970 British Cohort Study: Twenty-Nine-Year Follow-Up, 1999-2000.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2017d). Millennium Cohort Study: Fifth Survey, 2012.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2017a). Millennium Cohort Study: First Survey, 2001-2003.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2017c). Millennium Cohort Study: Sixth Survey, 2015.

University Of London. Institute Of Education. Centre For Longitudinal Studies (2017b). Millennium Cohort Study: Third Survey, 2006.

van de Schoot, R., P. Lugtig, and J. Hox (2012, July). A checklist for testing measurement invariance. *European Journal of Developmental Psychology 9*(4), 486–492.

Vandenberg, R. J. and C. E. Lance (2000, January). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods 3*(1), 4–70.

Vandenberg, R. J. and N. A. Morelli (2016). A contemporay update on testing for measurement equivalence and invariance. In *Handbook of Employee Commitment*. Edward Elgar Publishing.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika 41*(3), 321–327.

Venables, P. H., R. P. Fletcher, J. C. Dalais, D. A. Mitchell, F. Schulsinger, and S. A. Mednick (1983, April). Factor structure of the Rutter 'Children's Behaviour Questionnaire' in a primary school population in a developing country. *Journal of Child Psychology and Psychiatry 24*(2), 213–222.

Verhoeff, F. H., P. Milligan, B. J. Brabin, S. Mlanga, and V. Nakoma (1997, December). Gestational age assessment by nurses in a developing country using the Ballard method, external criteria only. *Annals of Tropical Paediatrics 17*(4), 333–342.

Victora, C. G., M. de Onis, P. C. Hallal, M. Blossner, and R. Shrimpton (2010). Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. *PEDIATRICS 125*(3), e473–e480.

Visram, A., P. Jasper, P. Vargas, J. Hug, and E. Jones (2017). Quantitative Impact Evaluation of the WINNN Programme. Technical report, Oxford Policy Management.

Wachs, T. D., M. Georgieff, S. Cusick, and B. S. McEwen (2014). Issues in the timing of integrated early interventions: Contributions from nutrition, neuroscience, and psychological research: Timing of integrated early interventions. *Annals of the New York Academy of Sciences 1308*(1), 89–106.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The lancet 366*(9499), 1804–1807.

Walker, S. P., S. M. Chang, M. Vera-Hernandez, and S. Grantham-McGregor (2011). Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *PEDIATRICS 127*(5), 849–857.

Wareham, N. J., E. M. F. van Sluijs, and U. Ekelund (2005, May). Physical activity and obesity prevention: A review of the current evidence. *Proceedings of the Nutrition Society 64*(02), 229–247.

Watkins, K. and M. Quattri (2016). Child poverty, inequality and demography. Technical report, Overseas Development Institute, London, UK.

Wen, S. W., R. L. Goldenberg, G. R. Cutter, H. J. Hoffman, S. P. Cliver, R. O. Davis, and M. B. DuBard (1990). Smoking, maternal age, fetal growth, and gestational age at delivery. *American Journal of Obstetrics and Gynecology 162*(1), 53–58.

WHO (2008). *Indicators for assessing infant and young child feeding practices*. Washington, D.C.: World Health Organization. OCLC: 795844943.

WHO (2009). *WHO Child Growth Standards: Growth Velocity Based on Weight, Length and Head Circumference: Methods and Development*. Geneva, Swtizerland: World Health Organization, Department of Nutrition for Health and Development. OCLC: ocn316014678.

Wiles, N. J., K. Northstone, P. Emmett, and G. Lewis (2009, April). 'Junk food' diet and childhood behavioural problems: Results from the ALSPAC cohort. *European Journal of Clinical Nutrition 63*(4), 491–498.

Wooldridge, J. M. (2014, September). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics 182*(1), 226–234.

World Bank (2017). PPP conversion factor, GDP (LCU per international USD). https://data.worldbank.org/indicator/PA.NUS.PPP.

Wu, H. and R. Estabrook (2016, December). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika 81*(4), 1014–1045.

Yoon, M. and R. E. Millsap (2007, July). Detecting Violations of Factorial Invariance Using Data-Based Specification Searches: A Monte Carlo Study. *Structural Equation Modeling: A Multidisciplinary Journal 14*(3), 435–463.

Zaman, S., R. N. Ashraf, and J. Martines (2008). Training in Complementary Feeding Counselling of Healthcare Workers and Its Influence on Maternal Behaviours and Child Growth: A Cluster- randomized Controlled Trial in Lahore, Pakistan. *Journal of Health and Population Nutrition 26*(2), 13.

Zhang, J., L. Shi, D.-f. Chen, J. Wang, and Y. Wang (2013). Effectiveness of an educational intervention to improve child feeding practices and growth in rural China: Updated results at 18 months of age: Child feeding intervention in rural China. *Maternal & Child Nutrition 9*(1), 118–129.