The health effects of early interventions: Evidence from Sure Start

Sarah Cattan* Gabriella Conti† Christine Farquharson‡ Rita Ginja¶

15 February 2018

Preliminary and incomplete - Do not cite without permission

Abstract

Early intervention programmes to deliver educational, social, and health services are widely recognised as a potentially effective way of promoting better health and preventing the development of costly health problems. While existing evidence demonstrates the potential benefits of targeted and often intensive interventions, we find that a fully universal, area-based intervention to bring together education, employment, and health services in England has benefits for childhood obesity and hospitalisations. We find that the introduction of Sure Start reduced the probability of being overweight at age 5 by one percentage point (10% of baseline), with stronger effects for white boys. Sure Start also reduced hospitalisations for respiratory illness and infections, consistent with predictions that exposure to additional pathogens would strengthen children's immune systems.

JEL Codes: I10, I14, I18

Keywords: Early intervention, obesity, hospitalisations, Sure Start

*Institute for Fiscal Studies. Email: sarah_c@ifs.org.uk.

†University College London, Department of Economics, Drayton House, 30 Gordon Street, London WC1H 0AX, United Kingdom. Tel.:+44(0)2076794696. Email: gabriella.conti@ucl.ac.uk; Institute for Fiscal Studies.

‡Institute for Fiscal Studies. Email: christine_f@ifs.org.uk.

§Department of Economics, University of Bergen, Fosswinckelsgate 14, 5007 Bergen, Norway. Tel.: +47 55 58 92 00. Email: rita.ginja@uib.no; Uppsala Center for Labor Studies (UCLS); IZA.

¶This paper is part of a project financially supported by the Nuffield Foundation, Project EYP 42289. The authors also acknowledge the excellent research assistant support of Teresa Steininger. We are grateful for the guidance of our advisory group: Richard Blundell (Institute for Fiscal Studies and University College London), Leon Feinstein (Office of the Children's Commissioner), Margaret Leopold (Department for Education), Joanne Pearmain (4Children), and Kathy Syvla (University of Oxford).

1 Introduction

Children in the earliest years of life are particularly sensitive to their environment, and are especially vulnerable to poor conditions that can have long-lasting impacts on their health (Shonkoff et al., 2009). However, interventions to improve health services (Bütikofer et al., 2015), promote healthy behaviours (e.g. Department of Health and Human Services, 2010), or foster supportive and nurturing environments have demonstrated benefits for health that can last well into middle age (D'Onise et al., 2010; Muennig, 2015). These interventions can be a cost-effective way of preventing disease, with benefits for both individual welfare and the public purse (Heckman et al., 2010).

While robust evidence comes from randomised controlled trials of small-scale intensive interventions in the U.S. (Campbell et al., 2014), recent research also finds health benefits from broader interventions such as universal preschool among disadvantaged populations (Carneiro and Ginja, 2014). However, much of the current evidence is based on targeted interventions in the U.S., with benefits frequently mediated through increased take-up of private health insurance (Conti et al., 2016; Muennig et al., 2009). Evaluations of universal interventions in Scandinavian countries, by contrast, have typically focused on long-run outcomes (e.g. Bütikofer et al., 2015).[1]

In this paper we present the first causal evidence of the health impacts of Sure Start, a major area-based early years initiative in England that combines parenting support, healthcare services, childcare, and employment support into 'one-stop shops' for children under the age of 5 and their families.

Exploiting variation induced by the 15-year rollout of the programme, we find evidence of a reduction in obesity for children, especially boys, who were more intensively exposed to Sure Start. Consistent with research from Scandinavia (Siflinger and van den Berg, 2016), we find evidence that Sure Start affects children's hospitalisations through effects on their immune system. In particular, hospital admissions for respiratory illnesses and infections rise in the early years of access to Sure Start as children are exposed to more pathogens, but subsequently fall as the child develops a stronger immune system more resistant to disease. These results indicate that area-based holistic early interventions can promote child health even in a country with universal, free healthcare and specific home-visiting services aimed at improving child health.

Remarkably, although Sure Start has been in place since 1999 its causal impacts on child outcomes have not yet been robustly evaluated. Our study adds to the literature in two main ways. First, our analysis complements two previous evaluations of the programme by employing a more robust econometric framework, using data on the exact location and opening date of each of the approximately 3,500 Sure Start facilities in England to obtain precise estimates of the causal impact of having access to Sure Start. We also extend previous work by using administrative data on hospitalisations and obesity to capture precise estimates of Sure Start's impact.[2] Second, more

[1]This is driven by the lack of data on outcomes of children rather than by choice, since the Scandinavia registry data used in this studies does not contain information on children on school performance before ages 14-15. An exception is a recent study by Siflinger and van den Berg, 2016, which uses health registry data.

[2]The timing of our study and the length of our data series also allow us to investigate longer-run effects than in

generally, the use of administrative health records since birth allow to study impacts during and after exposure to the program.

Sure Start was first evaluated by the National Evaluation of Sure Start (NESS) project, which collected data on children living in neighbourhoods served by the earliest Sure Start facilities. These children were compared to others surveyed in an earlier national survey who lived in areas not served by the earliest Sure Start programmes. The evaluation finds an increase in parent-reported hospitalisations at 9 months, no change in health outcomes at age 3, and lower Body Mass Index (BMI) and better parent-reported health status by age 5 for children living in the Sure Start neighbourhoods (NESS 2005, 2008, and 2010). However, the authors acknowledge that the non-random rollout of Sure Start, which was initially targeted at the most deprived areas, and the differences in timing and methodology of the outcome data between treated and control children are limitations of their approach.

Following NESS, Sure Start was next evaluated by the Evaluation of Children's Centres in England (ECCE). This study used a contextualised value-added model to analyse changes in health over the time that children and their families were enrolled in the study. They found no significant effects of visiting Sure Start centres on child health status, but some impact of using specific Sure Start services - health visitors and formal childcare - on the probability of changing health status (ECCE, 2015).

While NESS and ECCE assessed the implementation of Sure Start, the evidence base for the impact of similar programmes on child outcomes draws heavily on work done on the U.S. Head Start intervention, on which Sure Start was modelled. One of the longest-running and largest-scale early intervention programmes in the U.S., the intervention was comprehensively evaluated via an RCT in the Head Start Impact Study.[3] The results of the HSIS found short-term benefits for children's health status, health service use, and health insurance coverage, but these impacts had begun to fade out by the end of third grade (see Department of Health and Human Services - DHHS, 2010, 2012). However, the lacklustre results might be partly explained by failing to account for the substitution between different types of public services (Kline and Walters, 2016) or by substantial heterogeneity in the effectiveness of Head Start centres (Walters, 2015).

Subsequent studies have found short-run benefits for other health outcomes. Frisvold and Lumeng (2011) find that attending Head Start full-time decreases obesity among four- and five-year-old boys relative to part-time attendance. Ludwig and Miller (2007) exploit a cut-off in grant-writing assistance that affected counties' funding to deliver Head Start. They find a drop in mortality among five- to nine-year-olds for conditions that could plausibly be affected by Head Start. Recent studies also find evidence of medium- and long-term benefits for health. Carneiro and Ginja (2014) rely on a discontinuity on participation in Head Start driven by eligibility rules into

previous evaluations. We will add this in subsequent drafts.

[3]The Head Start Impact Study was the first experimental study of a large scale preschool program in the world. It finds that Head Start has short-term impacts on the cognitive and socio-emotional development of its participants, but that these disappear by first grade. While there are grounds on which this study can be criticised (e.g., Zigler (2010) and Kline and Walters (2016)), its main findings are notorious because of its transparent and rigorous design.

the programme and they find a reduction in the use of special equipment (eg, wheelchairs) among boys aged 12-13 and a reduction in obesity and the prevalence of depression among boys 16-17 years old. However, the small sample prevents the analysis of the impacts on the use of medical services.[4] In the longer run, Head Start has strong and lasting benefits for adult earnings, education, health insurance coverage, and risky behaviours such as smoking (Thompson, 2017; Muennig et al., 2009). These echo the long-run benefits of smaller, more intensive and targeted interventions such as the Carolina Abecedarian (ABC) and Perry Preschool Program (PPP) (e.g., Campbell et al., 2014; Conti et al., 2016), despite the much lighter-touch model adopted by Head Start. For example, Deming (2009) finds that the benefits of Head Start are around 80% as large as for ABC and PPP.

Although the evidence base for programmes such as Head Start, ABC, and PPP is rich, there is less evidence to understand how fully universal early intervention programmes affect child outcomes. This is particularly relevant since these programmes tend to be popular with European governments, where the institutional context typically offers a stronger social safety net. For example, mediation analysis for PPP finds that health insurance coverage is an important channel through which the programme acts. In countries with universal health insurance, this channel will not be available, which may attenuate some of the impact.

Available evidence from Europe suggests that universal early years interventions - particularly expanding access to pre-school education and formal childcare - can still benefit child development (e.g. Felfe and Lalive, forthcoming; Havnes and Mogstad, 2011; Black et al., 2014; by contrast Baker et al. (2008) find negative effects of universally subsidised childcare in Canada). Siflinger and van den Berg (2016) study a reform that provided universal childcare in one region of Sweden. Using hospital registry data for inpatient and outpatient visits by children aged 1 to 7, they find that decreasing the price for childcare at age 1 decreases the number of medical visits at ages 4-5 and 6-7. They interpret these findings as the result of an income effect, where families that pay less for childcare have more resources available for the consumption of other inputs into early child health, such as more nutritious foods.

Bhalotra et al. (2017) investigate an earlier Swedish programme: the introduction of universal post-natal health care, information, and support in the 1930s. Using a linked dataset of digitised birth records matched to mortality data, they find that the programme led to a 23% reduction in infant mortality and a further 6.5% increase in the probability of surviving to age 75. The authors note that the very long-run nature of the evaluation means that the contemporaneous policy context of the intervention was much different than today's, and in particular much more similar to the levels of social safety net provision found in

Bütikofer et al. (2015) also evaluate the very long-run impacts of a 1930s programme of mother and child health centres and post-natal home visiting in Norway. They use a similar empirical strategy to ours, linking administrative data to information on the rollout of the programme. They find that access to the well-child visits during the first year of life increased schooling and earnings.

[4]The authors do not find impacts of programme eligibility on the likelihood of dental check-ups in the 24 months preceding the survey at ages 6-7, 9-10 or 12-13; however, the sample size in estimations is about 75% of that used for the other outcomes presented in the paper.

They suggest that the centres could improve child nutrition through providing nutritional information (including breastfeeding promotion); improving dental health, and thereby the economic prospects of low-income children; and reducing the burden of other diseases, which can trigger an inflammatory response that undermines the benefits of proper nutrition. The authors find that the centres significantly reduced the incidence of obesity and cardiac events in men (though there were no effects for women), and that the effects were stronger for centres that offered a wider variety of health services.

Our paper complements these results by employing a similar strategy to look at more immediate child health outcomes. In common with much of the rest of the literature, we evaluate an early years programme that incorporates health services but is not solely targeted at improving health outcomes. Our results provide evidence that a fully universal intervention that is much less intensive than 'model' programmes like PPP and ABC can deliver benefits even in an institutional context with free at the point of use healthcare and pre-existing universal home-visiting services. Our focus on early health outcomes that are strongly linked to health status later in life complements and extends the existing evidence base for early intervention.

The paper proceeds as follows. Section 2 presents the institutional background of Sure Start and the services provided under the intervention. Section 3 discusses the different channels through which these services might influence obesity and hospitalisations. Section 4 details our analytical strategy and Section 5 the data we use. Section 6 presents our results and Section 7 concludes.

2 Institutional Background

Sure Start is one of the largest early childhood programmes in the United Kingdom, accounting for roughly one sixth of the early years budget in England in 2012-13 (Stewart and Obolenskaya, 2015). It was conceived as a network of centres bringing together new and existing services aimed at children under the age of five and their families. These centres were meant to be 'one-stop shops' where families could access programming and services targeting health, child development, parenting, and parental employment, as well as childcare services.

Sure Start was conceived as an area-based intervention whose services would be available to all families in the neighbourhood of the centre, with no means-testing (Melhuish, Belsky, and Leyland, 2010). When the first Sure Start Local Programmes (SSLPs) were opened in 1999, policymakers deliberately targeted more disadvantaged neighbourhoods with high levels of teenage pregnancy and low birth weight. Catchment areas were intended to be small ("perhaps one to two miles in radius in urban areas"), to be walkable, and to "make sense to the local community" (Department for Education and Employment - DfEE, 1999).

SSLPs were designed and encouraged to be responsive to local needs, and partnerships and programme managers had wide latitude in the services they offered. However, all SSLPs were required to deliver services that would meet key principles such as integrating services; involving parents; avoiding stigma; targeting specific objectives; and evaluating progress towards those goals

Table 1: The timeline of Sure Start

	Sure Start Milestones
Jan. 1999	First 60 'trailblazer' districts identified, invited to submit applications
Nov. 1999	Full approval of the first 15 Local Programmes
2000	Government target rises from 250 to 530 Local Programmes - Funding more than doubles
2003	Government pledges to universalise and expand SS Transition from Local Programmes to SSCCs
Dec. 2004	10-Year Strategy for Childcare pledges a SSCC in every community by 2010
2004-06	Phase One - targeting 20% most disadvantaged areas Most SSLPs transition to SSCCs
2006-08	Phase Two - targeting 30% most disadvantaged areas
2008-10	Phase Three - 'a Children's Centre in every community'

(DfEE, 1999). Sure Start was given a budget of £450 million over the period 1999-2002 to set up 250 projects, anticipated to reach 150,000 children over seven to ten years (Pugh and Duffy, 2010). The public reception of SSLPs was extremely positive.[5] Despite the opposition of the senior civil servants running the programme, in 2000 the government announced that it would more than double the programme target, from 250 SSLPs to 530 (Eisenstadt, 2011). In total, the government opened 524 Local Programmes.

In 2003, as part of a new strategy on the early years, the government announced that, rather than being phased out after 10 years, Sure Start would be universalised and supported indefinitely. Under the new strategy, the SSLPs and some other area-based early years services would transition into Sure Start Children's Centres (SSCCs), and responsibility for the programme would be transferred to local governments (Lewis (2011), p. 76). The following year, the 10-Year Strategy for Childcare pledged "a children's centre in every community" by 2010.

Although SSCCs were clear successors to the Local Programmes, they brought with them an increased focus on improving parental employment outcomes, both through integration with the new JobCentrePlus initiative and through offering free childcare under a national programme to improve parents' access to the labour market (Lewis, 2011). The Children's Centres also had less autonomy over the services they would deliver. Centres in the 30% most deprived areas were required to provide the "Core Offer", which consisted of integrated early education and childcare; parental outreach; family and parenting support; child and family health services; and links with JobCentrePlus (Lewis, 2011). To be "designated" as a Children's Centre, a centre in the most

[5] Director of the Sure Start Unit Naomi Eisenstadt reports that "The only complaint that [the government was] hearing from Members of Parliament about Sure Start was, 'Why can't my constituency have one?"' (Eisenstadt, 2011, p.48).

disadvantaged areas had to have, at a minimum, early years provision and links to JobCentrePlus in place. It then had up to two years to develop the remaining services or “full core offer”. For centres serving the 70% least disadvantaged areas, mandatory services were less intensive, but all centres had to offer activities for children and links to JobCentrePlus at designation, and were expected to develop health and outreach services. All centres were expected to open 5 days a week, 10 hours a day, 48 weeks a year.

Following the defeat of the New Labour government in 2010, the coalition Conservative-Liberal Democrat government focused on the existing network of 3,500 centres. They increased the focus on childcare and parental employment; removed the ring-fenced grants to fund Sure Start in favour of a broader, non-ringfenced Early Intervention Grant to local government; and reformed the “core offer” into a less prescriptive “core purpose”. This focused more on the outcomes the centres wanted to achieve for young children and their families, and gave LAs the freedom to develop Children’s Centres in a way which met local needs. The core purpose was incorporated into the revised statutory guidance for Children’s Centres published in April 2013.

3 Theoretical Framework

Sure Start was a holistic early years intervention, offering a diverse set of services and targeting improvement across a broad range of outcomes including child development, parental employment, and maternal and child health. In order to meaningfully interpret its effects on obesity and hospitalisations, we discuss the different channels through which it could affect outcomes.

Understanding the potential mechanisms is particularly important for interpreting the results on hospitalisations. While the clinical literature provides clear guidance on which levels of Body Mass Index are desirable, the interpretation of an increase in hospitalisations is more ambiguous. On the one hand, a child’s use of medical care can be interpreted as a positive outcome if families become more aware of and willing to use health services when needed, changing the selection into health service usage conditional on health status. Alternatively, increased use of health services could also reflect worse health.

The direct health service channel As mentioned in Section 2, one component of both the SSLPs and SSCCs was the direct provision of health services. Guidance on health service availability in Children’s Centres, for example, suggests that SSCCs should offer antenatal education; appropriate maternity services (including postnatal support); breastfeeding promotion and support; immunisation promotion; advice on accident and injury prevention; advice on obesity, diet, and nutrition; promotion of active play; support for mental health and for families with disabilities; speech and language development; and links to specific programmes, such as the Family Nurse Partnership (FNP) programme (DfE, 2010).[6] If these services increase the resources for or families’ access to direct health services in Sure Start areas, the presence and prevalence of Sure Start might

[6]The FNP is a preventive programme offering intensive, structured, nurse-led home visits for vulnerable first-time young mothers from early pregnancy until their children are two.

be expected to improve children's overall health status through a direct increase in health service provision.

In practice, a comprehensive analysis conducted in 2013 shows that 92 of the 121 centres evaluated by 2012 offered the services of a health visitor, but not direct medical services (DfE, 2013). This suggests that Sure Start had a relatively small effect on the *direct* provision of treatment-oriented health services within the UK; hence, the centres' users might have sought health care elsewhere, especially in case the provision of information in the SSCCs had increased awareness of their importance, as outlined in the next paragraph.

The information channel In addition to the provision of health services, Sure Start increased access to information about health. This included both support for healthy behaviours (such as breastfeeding) and referrals to health services. Although the DfE evaluation suggests that many Children's Centres provided relatively low levels of direct health services, screening and health visitations appear to have been major activities. The provision of such services would then be consistent with increased demand for medical care due to improved information, a greater willingness to use health services (for example, through lower stigma), and/or better screening for conditions that might benefit from treatment.

On the other hand, parents also received information on preventing illness and accidents in their children. To the extent that higher awareness reduced the incidence of health conditions and accidents, greater provision of information through Sure Start would also be consistent with a reduction in hospitalisations due to improved health status.

The household income channel While the direct services and the information channels pertain to health-related activities undertaken by Sure Start, there are numerous other inputs that can affect child health indirectly. Many of these inputs, such as nutritious food or high-quality childcare, have monetary costs associated with them. Family income therefore plays a role in the quality and quantity of these inputs that parents can provide to their children.

Sure Start can affect these decisions primarily through its focus on increasing family income through moving parents into work and providing additional job-related training. The effects of increasing parental labour force participation are ambiguous: increased hours of work will boost family income, increasing the monetary resources parents have to invest in their children. On the other hand, increased labour implies reduced leisure; parents who move into work (or take on more hours at their jobs) will have fewer hours at home with their children, who might therefore receive fewer time-intensive investments that are delivered by parents. For example, a mother who moves into full-time work when her child is aged 6 months might decide that continued breastfeeding is too time-consuming and inconvenient, but she might also be able to purchase higher-quality formula and baby food.

The exposure channel Another effect of using Sure Start services is to bring together children and parents from different families. Spending more time in close proximity to other children could increase a child's exposure to bacteria, viruses, and other pathogens to which they are not yet immune. This could in turn increase the number of illnesses contracted, which might lead to a rise in hospitalisations if the illnesses are serious.

Again, this rise in hospitalisations could be interpreted as a negative impact - more childhood illnesses and a worse state of health. On the other hand, increased exposure to a variety of pathogens early in life can boost a child's immune system, reducing readmissions or lessening the risks to their health at later ages (see for example Henderson et al. (1979) and Siflinger and van den Berg (2016)).

Expected effects on hospital admissions In sum, the effects of Sure Start on hospitalisations depend on a number of channels and are theoretically ambiguous. Although existing evidence that disentangles these different channels is scarce, there is some suggestion that increased information and higher family income will increase demand for medical services. The short-run effects on child health are ambiguous and depend on the exposure and information channels, but the long-term effect (and the effect on conditions like accidents where the exposure channel may matter less) should be a reduction in hospitalisations.

4 Empirical strategy

Our goal is to estimate the causal impact of having access to Sure Start on children's weight and obesity indicators and use of hospital services. To do so, we exploit the staggered introduction of Sure Start Local Programmes and Children's Centres (SSLPs and SSCCs, respectively) between 1999 and 2012. We use a difference-in-difference strategy to estimate the effect of increasing access to Sure Start, measured by an increase in the number of facilities per thousand children aged 0-4 in a local area.

The major challenge to this approach is that the rollout of Sure Start was not random. In the early part of the rollout policymakers explicitly targeted local authorities with higher levels of deprivation, teen pregnancy, and low birth weight; later, the universalisation of Sure Start prioritised more deprived areas to receive centres earlier.

We overcome this issue in two ways. First, we exploit the large size of our administrative datasets to include very low-level fixed effects at the level of the Lower Super Output Area (LSOA) to control for permanent unobserved differences across areas. There are 32,482 of these neighbourhoods in England, each containing around 1,500 residents.

However, LSOA fixed effects will not account for any time-varying local characteristics that could influence both the rollout of Sure Start and the health outcomes that we measure. To understand whether local level trends correlate with the rollout of SS we proceed in two ways. First, in our main estimating models we control explicitly for Local Authority linear trends. Second to study trends in local characteristics that potentially may correlate with an earlier adoption of SS, we construct a rich panel of local area characteristics, including both the traits that the rollout guidelines explicitly considered (ie, deprivation, low birth weight, and teenage pregnancy) and a set of other observable characteristics that might plausibly have influenced the rollout (including demographics, labour market statistics, pre-existing health and education services, and measures of political alignment).[7] Using this dataset, we analyse how both the levels and trends of observable traits are correlated with three dimensions of the rollout of Sure Start: whether a neighbourhood ever opened a Sure Start facility, when it opened, and the intensity of coverage in each local authority. We perform this analysis separately for SSLPs and SSCCs to capture the different goals of each phase of the programme. In this section, we summarise the key findings from this analysis and discuss their implications for our empirical strategy.

4.1 The rollout of Sure Start

The Sure Start Unit's guidance on the rollout was designed to ensure that areas fulfilling its criteria - deprived, with poor health indicators and a large population of children, spread across England and across different types of areas - would get Sure Start facilities, and would open them earlier and more intensively. In general, we find below that the rollout was faithful to these criteria, but also influenced by factors outside of guidelines.

Sure Start Local Programmes Table 2 summarises the results, indicating whether observable area characteristics were positively or negatively correlated with each of the three dimensions of the rollout. These results come from joint regressions including the predictors listed in Appendix A. The regressions are run at different levels of geography: the analysis of 'ever SSLP' is run at the electoral ward level, 'first SSLP' is run at the ward level for the subset of wards that did open an SSLP, and 'coverage SSLP' is run at the Local Authority District level to ensure that the variation is driven by different number of Sure Start facilities rather than differences in population.[8]

[7]Full details of the variables we include and the geographic level(s) at which they are measured are included in Appendix A.

[8]In 1998, there were 8,412 wards in England with an average population of about 6,000. There are 326 Local Authority Districts in England.

Table 2: Summary of significant predictors of the rollout of SSLPs

	Ever	Early	Coverage
1998 level/one-year lags			
Deprivation	+	+	+
Population (LAD)			
Pop 0-4 (LAD)			
Pop 0-4 (ward)	+		
density (LAD)		+	-
low birth weight		-	
teen conception rate			+
fertility rate		+	
infant mortality rate			
births		-	
coverage in region		-	+
GPs per thousand in LAD			+
ever family centre in ward	+		
EECs per thousand in LAD		-	
Children Looked After rate	+		
childcare age 3			
childcare age 4			
JSA rate			
employment rate			-
male FT wkly pay			+
female FT wkly pay			-
LAD aligned with nat'l		+	
LEA aligned with nat'l		+	
LAD and LEA aligned with nat'l		-	
LAD marginal			
LEA marginal			
4- or 5-year growth rates			
Population (LAD)			
Pop 0-4 (LAD)			
% low birth weight			
fertility rate			
total births			
GPs per thousand in LAD			
JSA rate			-
employment rate			

Notes: (1) LAD = Local Authority District; LEA = Local Education Authority (county).
(2) Deprivation is measured based on the 2000 Index of Multiple Deprivation.
(3) Teen conception rate = conceptions to women under 18 per thousand women aged 15-17; births = standardised total births.
(4) Coverage in the region is a standardised measure of the coverage in the same geographic region.
(5) GPs = General Practitioners; EECs = Early Excellence Centres, another area-based programme.
(6) Children Looked After rate is the share of children who have been taken into local authority care. 'Childcare' refers to the proportion of 3-/4-year-olds taking up a free childcare place.
(7) JSA = Jobseekers Allowance, the main unemployment benefit. Female and male earnings variables are median weekly pay among full-time employees.
(8) Political alignment variables measure whether the lower-level (LAD) and upper-level (LEA) councils were politically aligned with the national government. 'Marginal' variable capture whether the largest party in the council holds 45-55% of the seats.

Each of the guideline variables identified by the Sure Start Unit - deprivation, low birth weight, teenage conception rate, and regional spread - was a significant predictor of at least some dimensions of the rollout. Consistent with the intention for SSLPs to target the most disadvantaged neighbourhoods, a high level of deprivation is the most consistent predictor of greater Local Programme presence.

In addition to these guidelines variables, a number of other characteristics were significantly associated with the rollout decisions. These predictors ranged across all dimensions of observable characteristics that we considered (potential demand, health and vital statistics, services, early years provision, labour market, and political variables). Growth rates in the variables for which we have data available are almost never significant over and above the relevant levels.

Sure Start Children's Centres Reflecting the change in administrative geography and the shift for policymakers to consider more localised measures of neighbourhood disadvantage, the rollout analysis for 'ever SSCC' and 'first SSCC' is run at the level of the Lower Super Output Area (LSOA), a low-level geography with an average population of 1,500. The coverage regression is still run at the LAD level. An overview of these results is presented in Table 3.

Earlier SSCCs often grew out of pre-existing services like SSLPs, Early Excellence Centres (EECs), and Neighbourhood Nursery Initiatives. Accordingly, one of the strongest predictors of the SSCC rollout is whether a LSOA ever opened an SSLP. Having an SSLP within the LSOA was strongly associated with ever opening an SSCC, an earlier opening date, and more intensive coverage. As with the SSLPs, deprivation was another strong and consistent predictor of all three dimensions of the SSCC rollout. This suggests that, while Children's Centres were meant to be spread across the country, policymakers treated more deprived areas earlier and more intensively than their better-off neighbours.

Districts in regions with a relatively high level of coverage the year before tended to have higher coverage, suggesting a degree of persistence in coverage. This also applies to coverage in similarly-dense areas, an indicator of whether Children's Centres are spread across urban and rural areas. Lower relative coverage in similarly-dense areas predicts an earlier opening date as policymakers tried to reduce inequality in service provision across different density types.

Higher female unemployment rates were associated with both less chance of having a Sure Start facility and a later opening date. This could reflect the focus of SSCCs on providing links with childcare services, which are particularly valuable to working mothers.

Table 3: Summary of significant predictors of the rollout of SSCCs

	Ever	**Early**	**Coverage**
2002 level/one-year lags			
SSLP in LSOA	+	+	+
Deprivation	+	+	+
Pop 0-4 (LAD)			-
Pop 0-4 (LSOA)	+		
density (LAD)	-		-
density (LSOA)	-	+	
low birth weight			
teen conception rate			
fertility rate			-
infant mortality rate		+	
SS coverage in region			+
SS coverage in similar areas		-	+
GPs per thousand in LAD			
JCPs per thousand in LAD			
% English as Additional Language		-	+
Unemp rate, women	-	-	
Unemp rate, men			
Unemp rate, low-ed women			
Unemp rate, low-ed men		+	
male FT wkly pay			
female FT wkly pay			
LAD aligned with nat'l		+	
LEA aligned with nat'l			+
LAD and LEA aligned with nat'l		+	
LAD marginal			
LEA marginal		-	

Notes: (1) LAD = Local Authority District; LEA = Local Education Authority (county); LSOA = Lower Super Output Area. (2) Deprivation is measured based on the 2000 Index of Multiple Deprivation. (3) Teen conception rate = conceptions to women under 18 per thousand women aged 15-17; births = standardised total births. (4) Coverage in the region is a standardised measure of the coverage in the same geographic region. Coverage in similar areas splits LSOAs into three types based on their density. (5) GPs = General Practitioners; JCP = JobcentrePlus, an employment centre. (6) English as Additional Language = % of 5- to 11-year-olds with a non-English native tongue. (7) Low-educated = left school at age 16 or before. Female and male earnings variables are median weekly pay among full-time employees. (8) Political alignment variables measure whether the lower-level (LAD) and upper-level (LEA) councils were politically aligned with the national government. 'Marginal' variable capture whether the largest party in the council holds 45-55% of the seats.

Political considerations were more significant for the Children's Centre rollout than for the SSLPs, with a particularly strong relationship to the year of opening: LSOAs in local authorities aligned with the national government (Labour from 1997-2010, Conservative and Liberal Democrat thereafter) tended to open centres earlier and to have higher coverage.

Overall, this analysis suggests that the official guideline variables for both the Local Programmes and the Children's Centres were among the most important determinants of the rollout, its timing, and its intensity. However, a number of additional characteristics, such as the potential demand for services (proxied by population and density), service provision, labour market characteristics, and political control, also influenced some dimensions of the rollout.

4.2 Econometric specification

As our analysis of the rollout process shows, Sure Start facilities opened first and in greater numbers in more deprived areas and in areas with more children aged 0-4. This means that a naive comparison of the outcomes of children living in areas with low vs. high coverage at one point in time would confound the impacts of Sure Start with those of other factors correlated with the type of area these two groups of children live in.

Instead, we propose to estimate the impact of SS by exploiting within-area changes in coverage across cohorts. Effectively, our estimator measures the change in health outcomes of cohorts born earlier (who had less SS access) with those of cohorts born later (who had greater access), within each area and controlling for secular trends in outcomes across cohorts.

Weight-related outcomes To look at child obesity and Body Mass Index (BMI), we estimate a model on individual-level data on the universe of Reception Year students (age 5-6) in 2007-2014. Thus, using the NCMP we are able to study impacts at age 5 for children born between 2003 and 2009. This implies that children identified attending the Reception Year could have potentially attended a SSCC (except for children born in 2002 who could have attended a SSLP). As the NCMP does not systematic collects information on the child's residence, we proxy the residence with the location of the school.

Our specification considers the impact of changing the coverage of SS $\delta CovSS_{dt}$ for children in a given LSOA. That is, for child i born in year t who attends school (a proxy for home address as this is not available in the data) in LSOA l located in local authority district d we estimate the following model:

$$y_{ildt} = \alpha + \delta CovSS_{dt} + L_l + C_t + \beta X_i + \gamma A_{l,d,t} + \rho t_{dt} + \epsilon_{ildt} \quad (1)$$

where $CovSS_{dt}$ is the ratio number of SS vacancies to children in a given age range at local authority district level. The model also includes fixed effects for the LSOA of the school (L_l); a set of cohort dummies C_t to control for systematic changes in BMI across birth-year cohorts; individual characteristics (ethnicity and and gender) X_i; time-varying local area characteristics $A_{l,d,t}$; and a set of linear time trends $trend_{d,t,p}$ for each LSOA-specific deprivation level p (low, medium, or high) interacted with each local authority district d.

We study different outcomes during the Reception year: BMI, probability of being obese (that is, BMI at percentile 95th or above for age), overweight (that is, BMI between at percentile 85 or above and below percentile 95 for age), normal weight (defined as BMI above percentile 2 and below percentile 85 for age), and, finally, underweight (defined as a BMI at percentile 2 or below for age).

We analyze heterogeneity in impacts under several dimensions. We allow impacts to vary by three broad levels of poverty of the area of residence (based on the Index of Multiple Deprivation), by urban vs. rural areas, by gender and by ethnicity of the child (white vs. non-white children). We further allow for interactions across these four groups (for example, effects for non-white boys residing in poorest 30% of LSOAs).

Hospitalisation outcomes We estimate a similar model to examine hospitalisations in the Hospital Episode Statistics (HES) data. Our analysis is performed on cells defined by LSOA of the home l, birth year t, age of admission a, and sex s.

While weight-related outcomes are all measured at one age after potential Sure Start treatment has ended, our analysis of hospitalisations considers outcomes for ages 1-5, when children are still undergoing treatment. We also consider admissions for different causes k. Separately for each admission age a and admission cause k, we estimate:

$$y_{ldts}^{ak} = \alpha + \delta^{ak} CovSS_{dt}^{a} + \lambda^{ak} Pop_{ldts}^{a} + L_l^{ak} + C_t^{ak} + \beta^{ak} X_{ldt}^{a} + \gamma^{ak} A_{ldt} + \rho^{ak} trend_{dtp} + \epsilon_{ldts}^{ak} \quad (2)$$

As before, $CovSS_{dt}^{a}$ is our main measure of SS treatment defined for each local authority district d and birth year t. In HES, it is also indexed by the outcome age a because we consider outcomes that occur prior to the end of treatment, and adjust the average coverage variable accordingly to ensure that it only relates to treatment that has actually been experienced at the time of hospitalisation. We control for the population of children in each LSOA-birth year-admission age-sex cell, Pop^a_{ldts}; fixed effects by LSOA (L^{ak}_l) and cohort (C^{ak}_t); time-varying local characteristics measured in the year of birth (A_{ldt}); time-varying local characteristics measured in the year of admission (X^a_{lt}); and the same set of linear time trends specific to each local authority district and LSOA deprivation level.

Our main identification assumption is that the error term in each specification ($\epsilon_{i,l,d}$ or $\epsilon^{a,k}_{l,t,s}$) is uncorrelated with the treatment variables $CovSS$, conditional on the covariates. In other words, the effect of the programme will be unbiased estimated if the opening of Sure Start centres is unrelated to any trends in the area-level characteristics.

We report results for four outcomes: admissions for any cause, admission for respiratory infections, admission for infectious and parasitic diseases, and admissions for external causes. We consider five outcome ages (from age 1-5) and extend our work by allowing for nonlinearities in the impact of Sure Start and for different impacts by gender, which has been an important type of heterogeneity in previous work (e.g. Conti, Heckman, and Pinto (2014)).

For all specifications, we present robust standard errors clustered at the level of local authority at the time of admission. Since the local authority of residence at the age of potential treatment is not available in the data, we must use the local authority of residence at admission.

5 Data

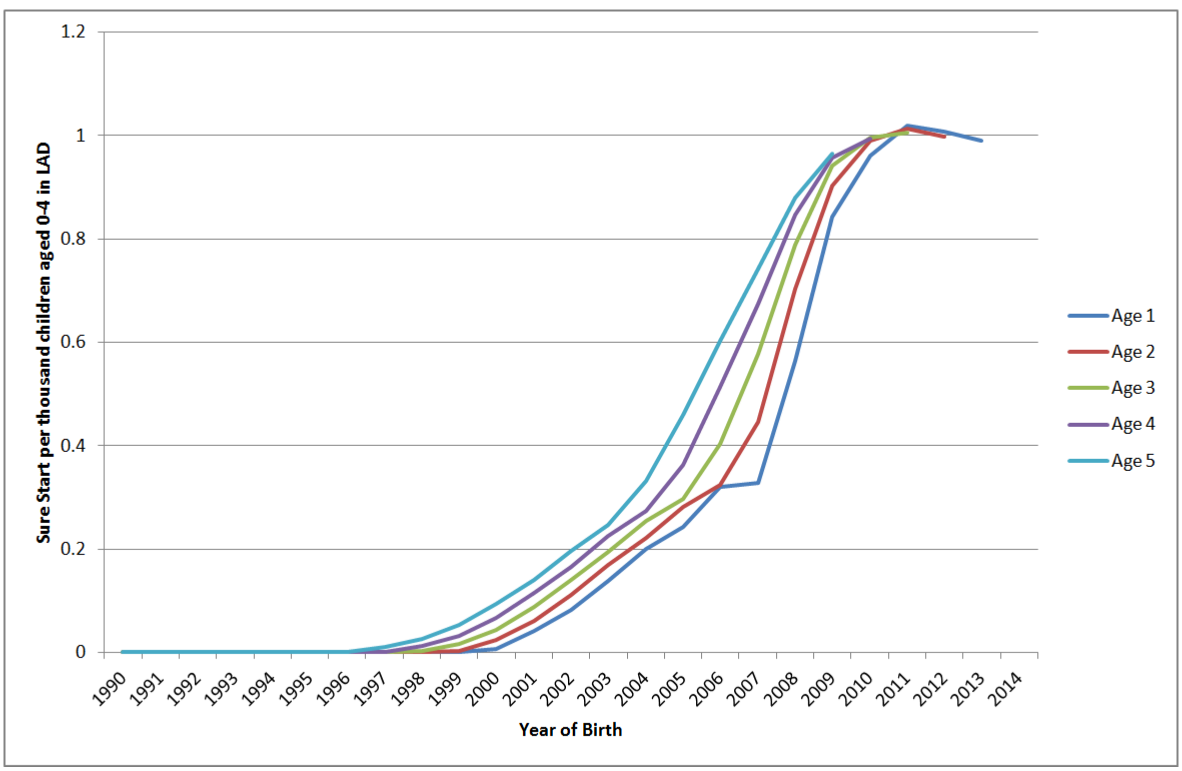
5.1 Data sources

Sure Start data Our main source of information about Sure Start is a dataset from the Department for Education (DfE) containing the postcode and date of opening of each Sure Start Local Programme and Children's Centre. We map the postcodes to their local authority district (LAD) to construct a panel of Sure Start coverage in each LAD, defined as number of centres per thousand children aged 0-4.

In constructing the level of coverage, we assume that Sure Start only affects children when they are between the ages of 0 and 4. Because of the challenges that centres faced in delivering services from their date of opening (Bouchal and Norris, 2014), we also allow centres to contribute to Sure Start coverage only from the start of the calendar year after they open. This means that the first cohort to be affected by a SSLP was born in 1996, and this exposure was only for one year when the children were aged 4. Children born in 1997 were affected for two years (at ages 3 and 4), and so on.

We pool information on SSLPs and SSCCs to create a harmonised measure of the coverage of Sure Start facilities. While the opening dates of all centres are precisely known, pooling in this way requires us to make an assumption about the closure dates of SSLPs. Since over 90% of SSLPs had transitioned into Children's Centres by 2006 (National Audit Office 2006, p.9), we assume that (a) any Local Programme that shares a postcode with a Children's Centre closed at the same time as the associated Children's Centre opened; and (b) all other Local Programmes closed at the end of 2006. It is therefore possible that we overstate the intensity of treatment between 2004 (when the first SSCCs were opened) and 2006, though the smooth profile of coverage across cohorts shown in Figure 1 suggests that this potential measurement error does not introduce major artificial discontinuities beyond age 1. The figure also shows that there is substantial variation in this measure of access; the coverage of Sure Start increased steeply throughout the 2000s, from no SS facilities in an LA amongst children born in 1995 to having on average 1 SS facility per thousand children in the LA amongst all cohorts born after 2010.

Figure 1: Average Sure Start coverage experienced, by cohort and age at hospitalisation



Note: Average coverage for each patient is a simple average of annual coverage between age 0 and the age prior to admission. The means shown here further average across all LSOA-sex cells for each admission age and birth year.

In our main results, we use a simple average of Sure Start coverage between age 0 and the age prior to outcome measurement to capture a measure of lifetime exposure to Sure Start. We use treatment up to the age prior to outcome measurement to ensure that all of the 'treatment' we consider has actually been experienced by the child. However, as illustrated in Table 4, this means that we ignore some treatment that did occur before outcome measurement.[9] Our approach is conservative and likely to understate the effect of Sure Start.

Table 4: Example of actual vs. measured treatment for children born in 2000 with outcomes measured at age 5

Born	**Outcome measure**	**Treatment**	**Coverage**					
			age 0	**age 1**	**age 2**	**age 3**	**age 4**	**age 5**
Jan/00	Jan/05	Actual	Jan/00 to Dec/00	Jan/01 to Dec/01	Jan/02 to Dec/02	Jan/03 to Dec/03	Jan/04 to Dec/04	Jan/05
Jan/00	Dec/05	Actual	Jan/00 to Dec/00	Jan/01 to Dec/01	Jan/02 to Dec/02	Jan/03 to Dec/03	Jan/04 to Dec/04	Jan/05 to Dec/05
Dec/00	Dec/05	Actual	Dec/00 to Nov/01	Dec/01 to Nov/02	Dec/02 to Nov/03	Dec/03 to Nov/04	Dec/04 to Nov/05	Dec/05
Dec/00	Nov/06	Actual	Dec/00 to Nov/01	Dec/01 to Nov/02	Dec/02 to Nov/03	Dec/03 to Nov/04	Dec/04 to Nov/05	Dec/05 to Nov/06
2000	Age 5	Measured	up to Dec/99	Jan/00 to Dec/00	Jan/01 to Dec/01	Jan/02 to Dec/02	Jan/03 to Dec/03	Jan/04 to Dec/04

Note: This table shows how actual treatment differs from measured treatment for four examples of children born in the same year (2000) and with outcomes measured at the same age (age 5). For age 5 outcomes, we construct an average treatment variable by averaging across the measured treatment at ages 0-4 (one year prior to the outcome age).

As our analysis of the rollout process shows, the timing of opening and the location of new Sure Start facilities were not random. In particular, Sure Start facilities opened first and in greater numbers in more deprived areas and in areas with more children aged 0-4. This means that a naive comparison of the outcomes of children living in areas with low vs. high coverage at one point in time would confound the impacts of Sure Start with those of other factors correlated with the type of area these two groups of children live in.

Instead, we propose to estimate the impact of SS by exploiting within-area changes in coverage across cohorts. Effectively, our estimator measures the change in health outcomes of cohorts born earlier (who had less SS access) with those of cohorts born later (who had greater access), within

[9] In particular, the decision to consider treatment up to the age prior to outcome measurement interacts with both our decision to define treatment at age A based on centres open in the year before the child turns A, and the error introduced by working in full calendar years to produce some measurement error in our treatment variable. We have chosen the most conservative approach in using these variables, but we are investigating the feasibility of using more precise data at the monthly rather than annual level to ameliorate one of these sources of error.

each area and controlling for secular trends in health outcomes across cohorts as well as time-varying area characteristics described below. We present our econometric specification in detail in the next sub-section.

National Child Measurement Programme (NCMP) The NCMP is an annual programme which measures since 2006 the height and weight of children in Reception and Year 6 within state maintained schools. In this draft we study outcomes up to ages 5-6, thus this data is used to estimate the impact of SS on height, weight, BMI, overweight and obesity for children in the Reception Year.

Participation in the programme is not compulsory, but non-participation is on an opt-out basis only. Before the programme starts each school year, Local Authorities write to the parents and carers of all children eligible for measurement to inform them of the programme and to give them the opportunity to opt their children out. The combination of this opt out possibility and to the fact that the programme was launched in 2006, there is a high and non-random rate of missing information in the data for the academic year of 2006-2007, for this reason we were advised by analysts at Public Health England, who is the entity responsible for the data management, to exclude this year from our analysis.[10]

Hospital Episode Statistics To study the impact on hospitalisations, we use the Hospital Episode Statistics (HES), an administrative dataset tracking the universe of patients using public English hospitals. Data on inpatient admissions has been collected since 1997-98 and tracks all hospital admissions, providing information on the admission, discharge, clinical diagnoses (up to 20 for each patient), and demographics of each patient. We include one record per hospitalisation (though patients may have several 'episodes' under different physicians during a single spell of hospitalisation) and exclude admissions related to the birth of a child.

Information on patient demographics is limited to sex, ethnicity, and month and year of birth; crucially, however, we also know their Lower Super Output Area (LSOA) of residence at the time of admission. This allows us to construct LSOA-level counts of all-cause and cause-specific admissions; to merge in Sure Start treatment variables (under the assumption of no mobility); to control for local authority trends; and to control for the time-varying area characteristics found to be important in the rollout analysis. To estimate the econometric model described in the previous section,

[10]See http://content.digital.nhs.uk/ncmp for information about the data.

we collapse the patient-episode level data to cells for each LSOA-birth year-admission age-sex combination, and fill the data so that all potential cells are represented.

In order to ensure comparability across outcome ages, we restrict our results on hospitalisations to a set of common cohorts, with only children born between 1996 and 2008 included in the regression. The lower bound is determined by the need to observe age 1 admission data in 1997 or later; the upper bound reflects our need to observe treatment up to age 7 in order to include leads of treatment as placebo leads. Our placebo testing strategy is described in more detail in section 6.3. To fix ideas, Figure 2 shows the cohorts that are used in the analysis by year of admission.

We report results for four outcomes: admissions for any cause, admission for respiratory infections, admission for infectious and parasitic diseases, and admissions for external causes. Based on the extremely skewed distributions and the high proportion of zeroes (see section 5.2 for details), we run linear probability models on binary outcomes (whether there was any admission in the cell). We consider five outcome ages (from age 1-5) and extend our work by allowing for nonlinearities in the impact of Sure Start and for different impacts by gender, which has been an important type of heterogeneity in previous work (e.g. Conti, Heckman, and Pinto (2014)). For all specifications, we present cluster-robust standard errors to account for an arbitrary covariance structure between the error terms of all cohorts living within each LAD at the time of admission.

Area-level characteristics Our analysis of the rollout of Sure Start suggests that it was correlated with a set of area-characteristics, and more importantly for our identification, with changes over time in these characteristics. Our preferred specifications outlined above aim to control for this potential source of bias by including time trends specific to each LAD and level of LSOA deprivation, and by including time-varying area characteristics as additional controls.

Specifically, we control for local area characteristics measured one year before birth, which include: population aged 0-4 (in the LAD and LSOA); population density; total period fertility rate; proportion of births below 2.5kg; number of GPs per thousand population; number of JobCentres-Plus per thousand population; Jobseekers' Allowance receipt rates (a proxy for unemployment); the proportion of three- and four-year olds taking up free childcare places; and political variables for district and county councils (including whether they are controlled by the same party in power nationally, and whether they are marginal, with the national government's party controlling between 45 and 55 per cent of the council's seats). We also control for five-year growth rates in population aged 0-4 and total population (of both the LSOA and LAD), as well as in the number of GPs per

Figure 2: Cohorts used in hospitalisations analysis, by year of admission

	Birth Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Pre-treatment	1990	Age 7																	
Pre-treatment	1991	Age 6																	
Pre-treatment	1992	Age 5																	
Pre-treatment	1993	Age 4	Age 5																
Pre-treatment	1994	Age 3	Age 4	Age 5															
SSLP only	1995	Age 2	Age 3	Age 4	Age 5														
SSLP only	1996	Age 1	Age 2	Age 3	Age 4	Age 5													
SSLP only	1997	*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5												
SSLP only	1998		*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5											
Both SSLP & SSCC	1999			*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5										
Both SSLP & SSCC	2000				*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5									
Both SSLP & SSCC	2001					*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5								
Both SSLP & SSCC	2002						*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5							
Both SSLP & SSCC	2003							*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5						
Both SSLP & SSCC	2004								*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5					
Both SSLP & SSCC	2005									*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5				
Both SSLP & SSCC	2006										*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5			
SSCC only	2007											*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5		
SSCC only	2008												*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5	
SSCC only	2009													*Age 0*	Age 1	Age 2	Age 3	Age 4	Age 5
SSCC only	2010														*Age 0*	Age 1	Age 2	Age 3	Age 4
SSCC only	2011															*Age 0*	Age 1	Age 2	Age 3
SSCC only	2012																*Age 0*	Age 1	Age 2
SSCC only	2013																	*Age 0*	Age 1
SSCC only	2014																		*Age 0*

Year of Admission

Years of Exposure: 1 2 3 4 5

Cohorts used in analysis

thousand.[11]

Finally, in the HES data, we control for several contemporaneous time-varying local area characteristics: the population count in each cell, the density of the LSOA, and the distance to the nearest hospital from the centroid of the LSOA. We also control for the share of admissions in each cell occurring in the spring, summer, and autumn.

5.2 Descriptive statistics

Body Mass Index and obesity Table 5 shows some basic characteristics for the weight of the universe of children enrolled in the Reception Year between 2007 and 2014. The table presents the mean BMI and the proportion of obese and overweight children by local deprivation, measured by deciles of the Index of Multiple Deprivation of the LSOA (Panel A) and by gender (Panel B). Panel A shows a strong economic gradient present in the mean of the three variables: children in the 10% more deprived LSOAs in England have on average a higher BMI, are more likely to be obese or overweight than those children living the 10% richest LSOAs: 12.1% of children in the Reception in the poorest LSOAs are obese, while this proportion reaches 6.8% in the richest LSOAs.

Table 5: Summary statistics for weight related outcomes

	(1)	(2)	(3)
	BMI	Obese	Overweight
Panel A: IMD Decile			
1	16.278	0.121	0.133
5	16.189	0.095	0.132
10	16.027	0.068	0.120
Panel B: Gender			
Girls	16.174	0.089	0.125
Boys	16.181	0.100	0.136

[11] Some of the variables that were present in the rollout analysis are excluded here. Median weekly earnings, the teen conception rate, and the infant mortality rate are all excluded because the series of data do not go back far enough. LAD-level unemployment rates cannot be used in this analysis because data security restrictions do not allow them to be exported from a secure environment, which cannot computationally support analysis on datasets as large as HES. We have attempted to proxy for them using the rates of receipt of Jobseekers Allowance, which is the chief unemployment benefit in the UK.

Table 6: Relative importance of selected causes of admission by age group

Condition	ICD-10 Codes		Age 1	Age 2	Age 3	Age 4	Age 5
Respiratory infections	J00-J99	% of admissions	31.1%	28.5%	28.6%	27.2%	25.4%
		% of cells >0	33.1%	23.5%	21.3%	19.4%	17.0%
Infectious and	A00-B99	% of admissions	23.3%	17.2%	12.9%	9.7%	7.9%
parasitic diseases		% of cells >0	27.1%	15.9%	11.0%	7.9%	5.9%
External causes	S00-T88, V00-Y99	% of admissions	10.9%	13.8%	13.3%	13.0%	13.5%
		% of cells >0	15.2%	13.8%	11.7%	10.7%	10.1%
All cause	All	% of admissions	100.0%	100.0%	100.0%	100.0%	100.0%
		% of cells >0	63.9%	53.0%	48.9%	47.1%	44.6%

Note: The numbers in this sample are based on the HES Inpatient Data for years 1997-2014. The rows '% of admissions' refer to the share of all admissions for each group accounted for by each condition. The rows '% of cells >0' refer to the proportion of LSOA-cohort-admission age-sex cells that have at least one admission for the relevant cause.

Hospitalisations We consider admissions both for all causes and for three of the most prevalent groups of conditions affecting young children: respiratory illness, infections and parasitic diseases, and external causes such as poisoning or broken bones. Table 6 summarises the ICD-10 codes and the importance of each condition as a proportion of total admissions, as well as the share of cells with at least one admission in each category.[12]

As shown in Table 7, these outcome variables are very non-normally distributed, with a high proportion of cells with no admissions and an extremely long right tail. Based on these distributions, we convert each outcome measure into a binary indicator for whether there is an admission in that category for each cell, and run a linear probability model on this indicator.[13]

6 Results

6.1 Obesity and BMI

Table 8 reports the estimates for δ in model 1. The table has four panels corresponding the variables studied: BMI and proportion of children obese, overweight and of normal weight. The

[12]When constructing measures of cause-specific admissions, we consider all diagnoses (including diagnoses that were not the primary reason for admission).

[13]We primarily rely on the LPM because computing constraints make it difficult to estimate non-linear models on over 800,000 observations while including over 32,000 fixed effects and 900 trends. However, we have repeated a few specifications using a logit, and the average marginal effects are broadly similar to the coefficients obtained from the LPM.

Table 7: Detailed summary statistics for hospitalisation outcomes

			Percentiles of Distribution				
Outcome	**Age**	**Mean (SD)**	**25th**	**50th**	**75th**	**95th**	**% positive**
Respiratory	1	0.536 (1.016)	0	0	1	2	33.1%
	2	0.342 (0.792)	0	0	0	2	23.5%
	3	0.300 (0.741)	0	0	0	2	21.3%
	4	0.263 (0.667)	0	0	0	1	19.4%
	5	0.222 (0.605)	0	0	0	1	17.0%
Infections & parasites	1	0.402 (0.821)	0	0	1	2	27.1%
	2	0.207 (0.562)	0	0	0	1	15.9%
	3	0.136 (0.451)	0	0	0	1	11.0%
	4	0.094 (0.380)	0	0	0	1	7.9%
	5	0.069 (0.336)	0	0	0	1	5.9%
External causes	1	0.188 (0.520)	0	0	0	1	15.2%
	2	0.166 (0.490)	0	0	0	1	13.8%
	3	0.139 (0.443)	0	0	0	1	11.7%
	4	0.126 (0.427)	0	0	0	1	10.7%
	5	0.118 (0.402)	0	0	0	1	10.1%
All causes	1	1.720 (2.631)	0	1	2	6	63.9%
	2	1.202 (2.405)	0	1	2	4	53.0%
	3	1.048 (2.320)	0	0	1	4	48.9%
	4	0.966 (2.217)	0	0	1	3	47.1%
	5	0.876 (2.083)	0	0	1	3	44.6%

Note: The numbers in this sample are based on the HES Inpatient Data for years 1997-2014 and reflect the distribution of the number of admissions in each LSOA-cohort-admission age-sex cell (including zeros). These include multiple admissions for the same child within a cell. '% positive' refers to the proportion of cells with at least one admission.

table includes also 7 columns: for the whole sample (column 1), for boys and girls (columns 2 and 3, respectively), and for each gender we also include in columns 4 to 7 the impacts by race.[14]

In column 1 of Table 8 an increase in the coverage of SS between ages 0-4 is not associated to a change in the BMI of children (Panel A). However, Panels C and D show that increasing the coverage of SS by 10 percentage points is associated to reduction in the proportion of overweight children by .12 percentage points (Panel C) with a corresponding increase in the proportion of normal weight children (Panel D). When we split the impacts by gender in columns 2 and 3, the decrease in the proportion of overweight children is entirely driven by boys. Furthermore, when in columns 4 to 7 the impacts by gender are studied separately by race (ie, white and non-white children), the reduction on the proportion of overweight boys is due to white boys. In future drafts, we will adjust inference to multiple hypotheses testing following the approach proposed by Romano and Wolf (2005).

As discussed in Section 4, the rollout of SS prioritized the poorest areas. Further, children in Reception Year in the NCMP were born between 2003 and 2009, and thus it is likely that the poorest areas in England were already served by a SS center. As the inception of the SS Children Centers in 2001, this poverty target was explicitly set at the 30% of the Index of Multiple Deprivation across LSOAs. Therefore, we expect, that increases in coverage of SS will have a stronger impact for children residing in LSOAs of middle income. To investigate whether this is the case, in Table 9 we include estimates for δ in model 1 by three levels of deprivation among LSOAs: poorest 30% LSOAs (Panel A), middle income LSOAs (Panel B) and richest 30% LSOAs (Panel C). Strikingly, the impacts are driven by children residing in middle income LSOAs, who given their cohorts of birth, would have been exposed to the sharpest increase in SS coverage.

[14]We do not report here the results on the proportion of underweight children since these represent a mere 1% of children enrolled in the Reception Year; the results are available from the authors and we are unable to detect any impact of SS on the proportion of underweight children.

Table 8: Impact of Sure Start on weight at Reception Year (age 5-6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Boys		Girls	
Sample	All	Boys	Girls	White	Non White	White	Non White
				Panel A: BMI			
Average	-0.044	-0.039	-0.055	-0.092*	0.072	-0.029	-0.153
coverage 0-4	(0.042)	(0.042)	(0.049)	(0.051)	(0.096)	(0.055)	(0.114)
Outcome mean	16.16	16.18	16.14	16.21	16.06	16.17	16.03
				Panel B: Obese			
Average	-0.002	-0.001	-0.002	-0.005	0.007	-0.001	-0.005
coverage 0-4	(0.004)	(0.005)	(0.005)	(0.006)	(0.015)	(0.007)	(0.014)
Outcome mean	0.096	0.104	0.088	0.098	0.122	0.083	0.108
				Panel C: Overweight			
Average	-0.012**	-0.018***	-0.006	-0.025***	-0.013	-0.002	-0.018
coverage 0-4	(0.005)	(0.006)	(0.007)	(0.008)	(0.012)	(0.008)	(0.016)
Outcome mean	0.129	0.136	0.122	0.142	0.113	0.128	0.107
				Panel D: Normal weight			
Average	0.012*	0.018**	0.006	0.027**	0.016	0.000	0.021
coverage 0-4	(0.007)	(0.008)	(0.009)	(0.011)	(0.018)	(0.010)	(0.022)
Outcome mean	0.762	0.746	0.779	0.751	0.734	0.782	0.765

Note: This table presents estimates for estimates for δ in model 1 using the NCMP data for the years of 2007-2014. Standard errors are clustered at the level of the LSOA. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

Table 9: Impact of SS on weight at Reception Year: Heterogeneity by deprivation of LSOA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	Panel A: Poorest 30% LSOAs			**Panel B: Mid-Income LSOAs**			**Panel C: Richest 30% LSOAs**		
	Panel A1: Obese			**Panel B1: Obese**			**Panel C1: Obese**		
Coverage 0-4	-0.001	0.008	-0.008	-0.005	-0.016**	0.008	-0.007	-0.006	-0.009
	(0.008)	(0.010)	(0.011)	(0.006)	(0.008)	(0.011)	(0.006)	(0.007)	(0.007)
Outcome mean	0.114	0.121	0.107	0.0976	0.106	0.0892	0.0804	0.0877	0.0726
	Panel A2: Overweight			**Panel B2: Overweight**			**Panel C2: Overweight**		
Coverage 0-4	0.002	-0.008	0.011	-0.022***	-0.036***	-0.008	-0.015*	-0.018*	-0.012
	(0.009)	(0.012)	(0.012)	(0.008)	(0.011)	(0.011)	(0.008)	(0.009)	(0.010)
Outcome mean	0.132	0.137	0.126	0.130	0.137	0.123	0.127	0.134	0.119
	Panel A3: Normal weight			**Panel B3: Normal weight**			**Panel A3: Normal weight**		
Coverage 0-4	-0.004	-0.004	-0.004	0.028***	0.055***	-0.001	0.018*	0.018	0.019
	(0.013)	(0.016)	(0.017)	(0.010)	(0.014)	(0.015)	(0.010)	(0.012)	(0.013)
Outcome mean	0.739	0.724	0.755	0.761	0.744	0.778	0.782	0.766	0.800

Note: This table presents estimates for estimates for δ in model 1 using the NCMP data for the years of 2007-2014. Standard errors are clustered at the level of the LSOA. *** Significant at 1%, ** Significant at 5%, * Significant at 10%.

The results presented in this version of the paper are robust to include additional controls for placebo coverage, that is, controlling by SS coverage at ages 7 to 9, at which ages children could no longer be in SS. These will added to future drafts.

6.2 Hospitalisations

Table 10 presents the results from model 2 when average Sure Start coverage enters linearly. The different panels show results for all-cause and cause-specific outcomes, and the dependent variable is an indicator for whether an LSOA-cohort-admission age-sex cell sees at least one admission for the relevant cause. Overall, we find little evidence of impact; while there is some suggestion that hospitalisations increase at younger ages, none of the coefficients is significant at the 5% level.

The evidence base from the U.S. stresses that early intervention programmes often have heterogeneous effects on outcomes by gender that can be masked when considering the outcomes of boys and girls together (e.g. Conti et al., 2016). Table 11 interacts the coverage variable with gender. We find much stronger evidence of an increase in hospitalisations among girls, with little effect on boys. This is in contrast to previous work, where males have tended to be more sensitive to early intervention programmes (e.g. Campbell et al., 2014; Carneiro and Ginja, 2014; Frisvold and Lumeng, 2014).

Although the effects on all-cause admissions are concentrated among girls, striking differences between genders emerge when considering respiratory and infection-related admissions. In Panel C, for example, cells containing the youngest boys are 3.5 percentage points more likely to have an admission when exposed to one additional unit of coverage (which roughly corresponds to the total increase in coverage over the 15-year rollout). This represents a more than 14% increase relative to the outcome mean. These effects fade out at older ages. The profile of effects for girls is similarly downward-sloping; however, rather than falling from positive to zero effect, Sure Start initially has little effect on girls, followed by strong downward pressure on admissions a few years later.

For both genders, the downward-sloping profile of coefficients corresponds to the exposure channel described in Section 3. At the youngest ages, children are exposed to new environments, people, and pathogens, which increases their chances of falling ill. However, these early illnesses also serve to strengthen the immune system, which after a few years is better able to fight off infections and illnesses. The lack of impact on admissions for external causes (which would not be affected by this immunisation pathway) is consistent with this explanation.

We also find evidence that the effects of Sure Start coverage on hospitalisations are non-linear.

Table 10: Effect of average Sure Start coverage on probability of any hospitalisation in a cell

	Admission age				
	1	**2**	**3**	**4**	**5**
Panel A: All-cause admissions					
Avg coverage	0.009*	0.010*	-0.003	0.007	0.009*
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
Outcome mean	0.637	0.533	0.492	0.477	0.453
Panel B: Respiratory illness					
Avg coverage	0.014*	0.008	0.002	0.002	0.009
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
Outcome mean	0.325	0.233	0.212	0.195	0.173
Panel C: Infections					
Avg coverage	0.014	0.006	-0.014*	-0.005	-0.006
	(0.009)	(0.009)	(0.008)	(0.007)	(0.006)
Outcome mean	0.246	0.148	0.107	0.081	0.064
Panel D: External causes					
Avg coverage	-0.001	0.001	-0.005	0.001	-0.005
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Outcome mean	0.149	0.137	0.117	0.108	0.103
N	853,800	853,800	853,800	853,800	853,800

Note: Data comes from the Hospital Episode Statistics and incorporates children born 1996-2008. Data is collapsed to the level of the LSOA-cohort-admission age-sex and outcomes are indicators for whether there was any admission of the specified type in the cell. Average coverage is an unweighted average of Sure Start coverage (centres per thousand children in the LAD) between age 0 and the age prior to admission.

Table 11: Effect of average Sure Start coverage on probability of any hospitalisation in a cell, by gender

	Admission age				
	1	**2**	**3**	**4**	**5**
Panel A: All-cause admissions					
Avg coverage, boys	0.003	0.007	-0.006	0.002	0.005
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
Avg coverage, girls	0.014***	0.013**	0.000	0.012**	0.013**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
Outcome mean	0.637	0.533	0.492	0.477	0.453
Panel B: Respiratory illness					
Avg coverage, boys	0.018**	0.012	0.011	0.008	0.016**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Avg coverage, girls	0.011	0.005	-0.007	-0.004	0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
Outcome mean	0.325	0.233	0.212	0.195	0.173
Panel C: Infections					
Avg coverage, boys	0.035***	0.024***	0.000	0.002	-0.001
	(0.010)	(0.009)	(0.008)	(0.007)	(0.006)
Avg coverage, girls	-0.006	-0.012	-0.027***	-0.012*	-0.012**
	(0.010)	(0.009)	(0.008)	(0.007)	(0.006)
Outcome mean	0.246	0.148	0.107	0.081	0.064
Panel D: External causes					
Avg coverage, boys	0.006	0.007	0.000	0.001	-0.003
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Avg coverage, girls	-0.008	-0.005	-0.010*	0.000	-0.007
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)
Outcome mean	0.149	0.137	0.117	0.108	0.103
N	853,800	853,800	853,800	853,800	853,800

Note: Data comes from the Hospital Episode Statistics and incorporates children born 1996-2008. Data is collapsed to the level of the LSOA-cohort-admission age-sex and outcomes are indicators for whether there was any admission of the specified type in the cell. Average coverage is an unweighted average of Sure Start coverage (centres per thousand children in the LAD) between age 0 and the age prior to admission.

Table 12 creates three bins of the average coverage treatment variable: no coverage (the omitted category), medium coverage (up to 0.3 centres per thousand children in the LAD), and high coverage (over 0.3 centres per thousand children). The cut-point between medium and high coverage corresponds to approximately the median of the distribution of cells with positive coverage at each age.

In general, the results of Table 12 suggest that there is some non-linearity in the effect of Sure Start coverage, with the effect of high coverage typically larger than that of medium coverage in the few cases where significant differences between them emerge. This is clearest in Panel C, where high coverage appears to be twice as influential as medium coverage in reducing hospitalisations for infections.

6.3 Placebo tests

For the results presented in the previous sections to be valid causal estimates of the impact of access to Sure Start, we require that the error term be uncorrelated with Sure Start coverage, conditional on the set of controls that we include. One potential threat to this would be if trends in the rollout of Sure Start are related to outcomes over and above the influence of the treatment itself, for example because of local area characteristics not accounted for in our control variables. To investigate this possibility, we run placebo tests by adding leads of future treatment to our specifications.

As outlined in Section 5, our efforts to ensure that the treatment variables that we interpret capture coverage that all children in the cell have experienced mean that the treatment variables we use do not capture all of the treatment experienced by some children. Put another way, some of the coverage that children have experienced prior to measuring their outcomes will included in treatment variables for later ages. As Table 4 indicates, this means that several of the leads of treatment are potentially 'contaminated', including information on treatment that some children in the cell will have experienced.

For this reason, for admissions at ages 1-4 the first valid placebo is our measured treatment at t+3. Because Sure Start is assumed to only treat children aged 0-4, in the age 5 specifications we can use treatment at age 7 as a clean placebo.[15] In HES, for example, our specification for admissions at age 1 includes treatment at age 0 (our estimated treatment effect), treatment at age

[15] A similar logic applies to the NCMP data, where we will use treatment at age 7 as a placebo to test our results. Unfortunately, these results are not yet available as we rely on the data owners to run the analysis on our behalf.

Table 12: Effect of bins of average Sure Start coverage on probability of any hospitalisation in a cell

	Admission age				
	1	**2**	**3**	**4**	5
Panel A: All-cause admissions					
Medium coverage	-0.001	-0.003*	-0.005***	-0.004**	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
High coverage	0.002	-0.004	-0.006***	-0.004	-0.004*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Outcome mean	0.637	0.533	0.492	0.477	0.453
Panel B: Respiratory illness					
Medium coverage	-0.002	-0.004*	-0.005**	-0.005**	-0.002
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
High coverage	0.002	-0.004	-0.005	-0.003	-0.006*
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Outcome mean	0.325	0.233	0.212	0.195	0.173
Panel C: Infections					
Medium coverage	-0.002	-0.002	-0.004**	-0.003**	-0.000
	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)
High coverage	-0.003	-0.004	-0.008***	-0.006**	-0.002
	(0.006)	(0.004)	(0.003)	(0.003)	(0.002)
Outcome mean	0.246	0.148	0.107	0.081	0.064
Panel D: External causes					
Medium coverage	-0.002	-0.002	-0.001	-0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
High coverage	-0.003	-0.004	-0.001	-0.000	-0.000
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Outcome mean	0.149	0.137	0.117	0.108	0.103
N	853,800	853,800	853,800	853,800	853,800

Note: Data comes from the Hospital Episode Statistics and incorporates children born 1996-2008. Data is collapsed to the level of the LSOA-cohort-admission age-sex and outcomes are indicators for whether there was any admission of the specified type in the cell. Bins of coverage are no coverage (omitted), medium coverage (up to 0.3 centres per thousand children in LAD), and high coverage (over 0.3 centres/1000). The bins are constructed from an average coverage variable that takes an unweighted average of measured treatment from age 0 to the year prior to admission.

Table 13: Effect of Sure Start coverage on probability of admissions, with placebo

	Admission age				
	1	**2**	**3**	**4**	**5**
Panel A: All-cause admissions					
Avg coverage	0.005	0.009	-0.003	0.003	0.007
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Placebo	-0.005	-0.004	-0.002	-0.008**	-0.005
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Panel B: Respiratory illness					
Avg coverage	0.016**	0.004	0.003	0.004	0.010
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Placebo	0.001	-0.005	-0.002	0.001	0.000
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)
Panel C: Infections and parasites					
Avg coverage	0.010	0.007	-0.014*	-0.008	-0.008
	(0.009)	(0.009)	(0.008)	(0.007)	(0.006)
Placebo	-0.000	-0.000	-0.002	-0.005	-0.001
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)
Panel D: External causes					
Avg coverage	-0.001	0.001	-0.004	-0.002	-0.009
	(0.005)	(0.007)	(0.006)	(0.005)	(0.006)
Placebo	-0.005	-0.004	-0.004	-0.003	-0.006*
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
N	853,800	853,800	853,800	853,800	853,800

Note: Data comes from the Hospital Episode Statistics and incorporates children born 1996-2008. Data is collapsed to the level of the LSOA-cohort-admission age-sex and outcomes are indicators for whether there was any admission of the specified type in the cell. Average coverage is an unweighted average of Sure Start coverage (centres per thousand children in the LAD) between age 0 and the age prior to admission. Placebos are treatment as measured at age t+3 for outcomes ages 0-4, and at age 7 for outcomes at age 5. Regressions also include measured treatment at all ages from t to t+2.

Table 14: Effect of bins of Sure Start coverage on probability of any hospitalisation in a cell, by gender and with placebo

	Admission age				
	1	**2**	**3**	**4**	**5**
Medium coverage, male	-0.001	-0.003	-0.003*	-0.002	-0.003
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
High coverage, male	-0.002	-0.005*	-0.006**	-0.005	-0.005*
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Medium coverage, female	-0.000	-0.002	-0.005**	-0.004**	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
High coverage, female	0.005	-0.000	-0.005*	-0.002	-0.001
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Medium cov, placebo male	0.003	0.002	-0.001	-0.005	0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
High cov, placebo male	0.004	0.004	-0.000	-0.007*	0.001
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Medium cov, placebo female	0.001	-0.001	-0.002	-0.003	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
High cov, placebo female	0.003	-0.001	-0.001	0.001	0.001
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
N	853,800	853,800	853,800	853,800	853,800
Joint p-value of placebos	0.808	0.614	0.940	0.107	0.267
Outcome mean	0.637	0.533	0.492	0.477	0.453

Note: Data comes from the Hospital Episode Statistics and incorporates children born 1996-2008. Data is collapsed to the level of the LSOA-cohort-admission age-sex and outcomes are indicators for whether there was any admission of the specified type in the cell. Bins of coverage are no coverage (omitted), medium coverage (up to 0.3 centres per thousand children in LAD), and high coverage (over 0.3 centres/1000). The bins are constructed from an average coverage variable that takes an unweighted average of measured treatment from age 0 to the year prior to admission. Placebos are treatment as measured at age t+3 for outcomes ages 0-4, and at age 7 for outcomes at age 5. Regressions also include measured treatment at all ages from t to t+2.

1 (a contemporaneous treatment measure which might not have been fully experienced, so we do not interpret it), treatment at ages 2 and 3 (which for some children in the cell will capture some of the treatment they have actually experienced), and finally treatment at age 4, which is a true placebo.

Table 13 shows the placebo tests for all outcomes using a linear specification, again finding that the placebo coefficients are not significant more often than would be expected by chance. However, including the placebo does attenuate some of the coefficients because the interim coefficients for treatment at t to t+2, which as discussed above contain some relevant information, are also added to the regression but not interpreted as treatment effects.

Table 14 shows that the placebo coefficients are also not significant (either singly or jointly) in a specification that incorporates both heterogeneity by gender and non-linear treatment effects.

7 Conclusion and next steps

In this paper we have presented the first causal evidence of the impact of Sure Start on child obesity and health service use. We exploit the staggered rollout of the programme over a 15-year period to run a difference-in-difference specification , accounting for the non-randomness of the rollout by incorporating local authority- and deprivation-specific time trends as well as a set of time-varying observable characteristics that we find to be predictive of the rollout.

We find that higher access to Sure Start reduces child obesity and promotes immune development, which is reflected in a short-term rise in hospitalisations followed by a decline in admission rates for respiratory illness and infections.

An increase in average coverage across the five years of treatment of one centre per thousand children in the local authority - equivalent to the increase in coverage experienced by the average local authority over the entire rollout of the programme - reduces the incidence of being overweight by 1.2 percentage points (10% of baseline) in the entire population of five-year-olds, with even stronger reductions of 2.5 percentage points for white boys. These effects are mirrored by an increase in the proportion of children at normal weight, emphasising the positive public health implications of these findings.

A similar increase in coverage has only weak effects on overall hospitalisations for any cause. However, extending the analysis to consider nonlinearities in Sure Start coverage suggests that higher levels of treatment reduce the probability of a hospital admission in an LSOA-cohort-sex-admission age cell by 0.5 percentage points at ages three and four. This is driven by impacts on respiratory illness and infections, while admissions for external causes such as accidents are unaffected. The overall effects are concentrated among boys, although the effect on admissions for particular causes is driven by reductions in admissions for girls.

While we use strong quasi-experimental variation to identify these effects, there are several limitations to our approach. First, we rely on measures of very local geography to define our treatment and incorporate fixed effects. However, our data sources only offer contemporaneous information about residence (HES) or school address (NCMP). This means that we must assume that any mobility in the child's first years of life is unrelated to outcomes (conditional on our covariates); this rules out responses such as spatial sorting. We minimise the potential for error by using addresses measured in the earliest years of life, but this is an important challenge for work on longer-term outcomes.

A second limitation of our approach is that we are not able to define the impact of actually using Sure Start services (rather than simply having greater access to them). However, our Intent to Treat estimates are still of substantial relevance for policymakers choosing whether and how to introduce similar area-based interventions without the ability to ensure take-up of the services themselves.

The policy implications of our work are clear: we find that universal, area-based, holistic interventions can be an effective complement to universal health care and other social services for improving child health and reducing obesity. Determining the short-run impacts that early interventions can have on health is important for a full cost-benefit analysis. The outcomes that we examine in this paper - child obesity and hospitalisations - are particularly salient for public policy. For example, the short-term health costs of treating child obesity in the UK are estimated at £51 million per year (Chief Medical Officer Annual Report, 2012). Our results suggest that government policy should explicitly consider early interventions that span a range of health and non-health services as part of an overall strategy of preventing illness and disease.

While our results present the first causal evidence on the health impacts of Sure Start, they are not the last word on this subject. Looking at both birth outcomes and hospitalisations at older ages in the HES data will offer a fuller understanding of how the impact of Sure Start evolves over the lifecycle, and whether the benefits we observe at younger ages persist. It will also be possible to look at the effect of treatment at different ages to understand whether Sure Start's services are particularly effective for one age group.

Finally, our analysis has focused on health outcomes found in administrative data. These have several advantages: they are well-measured, with a sufficiently large sample size to detect small effects, and are easily associated with public finance costs. However, Sure Start is a complex intervention that encompasses a range of services, and our results do not address which aspects of the programme are particularly beneficial for child health. Performing a mediation analysis to understand how Sure Start influences health behaviours such as nutrition or exercise would provide more insight into which aspects of the programme promote child health.

Our results confirm the important role that early interventions can play in preventing poor health. Based on the strengthening evidence base, policymakers should consider holistic early years programmes as part of an overall strategy to promote health, with benefits both for families and finance ministers.

References

[1] Bütikofer, Aline, Katrine Løken, and Kjell Salvanes. 2015. "Long-Term Consequences of Access to Well-Child Visit". *IZA Discussion Paper Series* 9546. Accessed 14 February 2018. http://legacy.iza.org/de/webcontent/publications/papers/viewAbstract??dp_id=9546.

[2] Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. "Universal Child Care, Maternal Labor Supply, and Family Well?Being". *Journal of Political Economy* 116 (4): 709?45. https://doi.org/10.1086/591908.

[3] Bhalotra, Sonia, Martin Karlsson, and Therese Nilsson. 2017. "Infant Health and Longevity: Evidence from A Historical Intervention in Sweden". *Journal of the European Economic Association* 15 (5): 1101?57. https://doi.org/10.1093/jeea/jvw028.

[4] Black, Sandra E., Paul J. Devereux, Katrine V. Løken, and Kjell G. Salvanes. 2014. "Care or Cash? The Effect of Child Care Subsidies on Student Performance". *The Review of Economics and Statistics* 96 (5): 824?37. https://doi.org/10.1162/REST_a_00439.

[5] Bouchal, Petr and Emma Norris. 2014. "Implementing Sure Start Children's Centres."In *Doing Them Justice: Lessons from four case studies of policy implementation*. London: Institute for Government.

[6] Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health". *Science* 343 (6178): 1478?85. https://doi.org/10.1126/science.1248429.

[7] Carneiro, Pedro, and Rita Ginja. 2014. "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start". *American Economic Journal: Economic Policy* 6 (4): 135?73. https://doi.org/10.1257/pol.6.4.135.

[8] Chief Medical Officer Annual Report. 2013. "Our Children Deserve Better: Prevention Pays". London: Williams Lea for Department of Health.

[9] Conti, Gabriella, James J. Heckman, and Rodrigo Pinto. 2016. "The effects of two influential early childhood interventions on health and healthy behaviors". *Economic Journal* 126(596): F28-F65. http://onlinelibrary.wiley.com/doi/10.1111/ecoj.12420/abstract

[10] Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start". *American Economic Journal: Applied Economics* 1(3): 111?34. https://doi.org/10.1257/app.1.3.111.

[11] Department for Education and Employment. 1999. *Sure Start: A Guide for Trailblazers*. London, UK: DfEE Publications.

[12] Department for Education. 2010. "Sure Start Children?s Centres: Statutory Guidance". London, UK.

[13] Department of Health and Human Services. 2010. "Head Start Impact Study: Final Report". Administration for Children and Families, Washington, DC.

[14] Department of Health and Human Services. 2012. "Third Grade Follow-up to the Head Start Impact Study: Final Report". Administration for Children and Families, Washington, DC.

[15] D'Onise, Katina, R. A. McDermott, and J. W. Lynch. 2010. "Does attendance at preschool affect adult health? A systematic review". Public Health 124(9): 500-511.

[16] Eisenstadt, Naomi. 2011. *Providing a Sure Start: How Government Discovered Early Childhood*. Bristol: Policy Press.

[17] Evaluation of Children?s Centres in England (ECCE). 2013. "Strand 3: Delivery of Family Services by Children?s Centres ? Research Report". Department for Education. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/224096/DFE-RR297.pdf

[18] Evaluation of Children?s Centres in England (ECCE). 2015. "The impact of Children?s Centres: Studying the effects of Children?s Centres in promoting better outcomes for young children and their families". Department for Education. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/485346/DFE-RR495_Evaluation_of_children_s_centres_in_England__the_impact_of_children_s_centres.pdf

[19] Felfe, Christina and Rafael Lalive. 2018. "Does Early Child Care Affect Children?s Development?" *CEPR Discussion Paper Series*, number DP12675. https://papers.ssrn.com/abstract=3118323.

[20] Frisvold, David E., and Julie C. Lumeng. 2011. “Expanding Exposure: Can Increasing the Daily Duration of Head Start Reduce Childhood Obesity?” *The Journal of Human Resources* 46(2): 373?402.

[21] Havnes, Tarjei and Magne Mogstad. 2011. “No Child Left Behind: Subsidized Child Care and Children?s Long-Run Outcomes”. *American Economic Journal: Economic Policy* 3(2): 97?129. https://doi.org/10.1257/pol.3.2.97.

[22] Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. “The Rate of Return to the High/Scope Perry Preschool Program”. *Journal of Public Economics* 94(1?2): 114?28.

[23] Henderson, Frederick W., Albert M. Collier, Wallace A. Clyde Jr, and Floyd W. Denny. 1979. “Respiratory-syncytial-virus infections, reinfections and immunity: a prospective, longitudinal study in young children”. *New England Journal of Medicine* 300(10): 530-534.

[24] Kline, Patrick, and Christopher R. Walters. 2016. “Evaluating Public Programs with Close Substitutes: The Case of Head Start”. *The Quarterly Journal of Economics* 131(4): 1795?1848. https://doi.org/10.1093/qje/qjw027.

[25] Lewis, Jane. 2011. “From Sure Start to children?s centres: An analysis of policy change in English early years programmes. *Journal of Social Policy* 40(1): 71-88. http://eprints.lse.ac.uk/32289/.

[26] Ludwig, Jens and Douglas L. Miller. 2007. “Does Head Start Improve Children?s Life Chances? Evidence from a Regression Discontinuity Design”. *The Quarterly Journal of Economics* 122(1): 159?208. https://doi.org/10.1162/qjec.122.1.159.

[27] Muennig, Peter, Lawrence Schweinhart, Jeanne Montie, and Matthew Neidell. 2009. “Effects of a Prekindergarten Educational Intervention on Adult Health: 37-Year Follow-Up Results of a Randomized Controlled Trial.” *American Journal of Public Health* 99(8): 1431?37. https://doi.org/10.2105/AJPH.2008.148353.

[28] Muennig, Peter. 2015. “Can universal pre-kindergarten programs improve population health and longevity? Mechanisms, evidence, and policy implications”. *Social Science and Medicine*, 127: 116-123.

[29] National Audit Office. 2006. "Sure Start Children?s Centres." London: The Stationery Office.

[30] National Evaluation of Sure Start (NESS). 2005. "Early impacts of Sure Start Local Programmes on Children and Families." Department for Education and Skills. http://www.ness.bbk.ac.uk/impact/documents/1183.pdf.

[31] National Evaluation of Sure Start (NESS). 2008. "The impact of Sure Start Local Programmes on Three Year Olds and Their Families." Department for Children, Schools, and Families. http://www.ness.bbk.ac.uk/impact/documents/41.pdf

[32] National Evaluation of Sure Start (NESS). 2010. "The impact of Sure Start Local Programmes on five year olds and their families." Department for Education. http://www.ness.bbk.ac.uk/impact/documents/RR067.pdf

[33] Pugh, Gillian and Bernadette Duffy. 2010. *Contemporary Issues in the Early Years*. London: Sage.

[34] Romano, Joseph P. and Michael Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping". *Econometrica* 73(4): 1237-1282.

[35] Shonkoff, Jack P., W. Thomas Boyce, and Bruce S. McEwen. 2009. "Neuroscience, molecular biology, and the childhood roots of health disparities: building a new framework for health promotion and disease prevention". *JAMA*, 301(21): 2252-2259.

[36] Siflinger, Bettina, and Gerard van den Berg. 2016. "The Effects of a Universal Child Care Reform on Child Health ? Evidence from Sweden". Annual Conference 2016 (Augsburg): Demographic Change. Verein fr Socialpolitik / German Economic Association. https://ideas.repec.org/p/zbw/vfsc16/145765.html.

[37] Stewart, Kitty and Pauline Obolenskaya. (2015). "The Coalition's Record on the Under-Fives: Policy, Spending, and Outcomes 2010-2015." *CASE Working Paper Series* number 12. London: Centre for Analysis of Social Exclusion.

[38] Thompson, Owen. 2017. "Head Start?s Long-Run Impact: Evidence from the Program?s Introduction". *Journal of Human Resources*, October, 0216?7735r1. https://doi.org/10.3368/jhr.53.4.0216.7735R1.

[39] Walters, Christopher R. 2015. “Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start”. American Economic Journal: *Applied Economics* 7 (4): 76?102. https://doi.org/10.1257/app.20140184.

[40] Zigler, Edward. (2010). “Putting the National Head Start Impact Study into a Proper Perspective”. National Head Start Association (U.S.). http://www.nhsa.org.

APPENDIX

A Rollout variables

Table 15 summarises the variables considered in each of the six specifications for the rollout analysis. The 'ever had' analyses are run on the set of wards/LSOAs (respectively for SSLPs and SSCCs); the 'first year of opening' analysis is run on the set of wards/LSOAs that ever opened an SSLP/SSCC; and the 'coverage' regressions measure coverage in all local authority districts across the period of the rollout.

Notes to Table 15:

(1) Y: The variable is included in the regression.

(2) 1998: The variables 1998 level is included in the regression.

(3) 2002: The variables 2002 level is included in the regression.

(4) t-1: The variables level in the year before the outcome measurement is included in the regression.

(5) Trend: The variables four-year growth rate to the year of measurement.

(6) Low education is defined as leaving school at or before age 16.

(7) LAD is the local authority district (lower-level). LEA is the Local Education Authority (upper-level).

Table 15: Variables included in the rollout analysis

	Ever had		1st year of opening		Coverage	
	SSLP	**SSCC**	**SSLP**	**SSCC**	**SSLP**	**SSCC**
Level of Analysis:	ward	LSOA	Ward	LSOA	LAD	LAD
Sure Start intensity						
Ever SSLP, LSOA		Y		Y		
Ever SSLP, LAD						Y
SSLP coverage, region			t-1		t-1	
SSCC coverage, region				t-1		t-1
SSLP coverage, density type		*does not exist - no ward area information available*				
SSCC coverage, density type				t-1		t-1
Deprivation						
linear rank of IMD	Y	Y	Y	Y	Y	Y
dummies of IMD categories	Y	Y	Y	Y	Y	Y
share of wards in IMD categories					Y	Y
Potential Demand						
population 0-4, LSOA		2002		t-1, trend		
population 0-4, ward	1998		t-1			
population 0-4, LAD	1998, trend	2002, trend	t-1	t-1, trend	t-1	t-1, trend
population, LAD	1998, trend		t-1		t-1	
Urbanicity						
density, LSOA		2002		t-1		
density, ward		*does not exist - no ward area information available*				
density, LAD	1998, trend	2002, trend	t-1	t-1, trend	t-1	t-1, trend
Health Guidelines						
% low birthweight, LAD	1998, trend	2002, trend	t-1	t-1, trend	t-1	t-1, trend
under-18 conception rate, LAD	1998	2002, trend	t-1	t-1, trend	t-1	t-1, trend
Vital Statistics						
total period fertility rate, LAD	1998, trend	2002, trend	t-1	t-1, trend	t-1	t-1, trend
infant mortality rate, LAD	1998	2002, trend	t-1	t-1, trend	t-1	t-1, trend
births, LAD	1998, trend		t-1		t-1	
Pre-existing services						
GPs per capita, LAD	1998, trend	2002, trend	t-1	t-1, trend	t-1	t-1, trend
JCPs per capita, LAD		2002		t-1		
ever family centre, ward	Y		Y		Y	
EECs per capita, LAD	1998		t-1		t-1	
free childcare take-up, age 3, LEA	1998		t-1		t-1	
free childcare take-up, age 4, LEA	1998		t-1		t-1	
Children Looked After rate, LEA	1998, trend		t-1		t-1	
Labour Market						
Jobseekers Allowance rate, LAD	1998, trend		t-1		t-1	
unemployment rate (women), LAD		2002, trend		t-1, trend		t-1, trend
unemployment rate (men), LAD		2002, trend		t-1, trend		t-1, trend
unemployment rate (women low-ed), LAD		2002, trend		t-1, trend		t-1, trend
unemployment rate (men low-ed), LAD		2002, trend		t-1, trend		t-1, trend
earnings (female weekly), LAD	1998	2002, trend	t-1	t-1, trend	t-1	t-1, trend
earnings (male weekly), LAD	1998	2002, trend	t-1	t-1, trend	t-1	t-1, trend
overall employment rate, LAD	1998, trend		t-1		t-1	
Child Language Development						
English as Additional Language, LAD		2002		t-1, trend		t-1, trend
Political Alignment						
Aligned with gov't, LAD	1998	2002	t-1	t-1	t-1	t-1
Aligned with gov't, LEA	1998	2002	t-1	t-1	t-1	t-1
Aligned with gov't, LAD and county	1998	2002	t-1	t-1	t-1	t-1
Marginal for gov't, LAD	1998	2002	t-1	t-1	t-1	t-1
Marginal for gov't, county	1998	2002	t-1	t-1	t-1	t-1