HCEO WORKING PAPER SERIES

The Health Effects of Universal Early Childhood Interventions: Evidence from Sure Start

Sarah Cattan
Gabriella Conti
Christine Farquharson
Rita Ginja
Maud Pecher

Working Paper

2021-051

11/2021; Revised 10/2022

HUMAN CAPITAL AND ECONOMIC OPPORTUNITY GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

The Health Effects of Universal Early Childhood Interventions: Evidence from Sure Start

Sarah Cattan, Gabriella Conti, Christine Farquharson, Rita Ginja and Maud Pecher*

October 12, 2022

Abstract

Early Childhood Interventions (ECI) offering disadvantaged children preschool and family support services in the US show long-lasting health impacts. Can these benefits hold when these programs are offered to all children in contexts with universal healthcare? We evaluate the short- and medium-term health impacts of Sure Start, a universal integrated ECI in England, exploiting its 11-year rollout and administrative hospitalizations data. One additional Sure Start center per thousand age-eligible children increases hospitalizations by 10% at age 1, but reduces them by 8-9% across ages 11-15. Impacts are concentrated in disadvantaged areas and likely driven by both health and non-health services.

JEL Codes: I10, I14, I18.
Keywords: early childhood intervention, health, difference-in-differences

*Cattan: Institute for Fiscal Studies and IZA (sarah_c@ifs.org.uk). Conti: University College London, Institute for Fiscal Studies, IZA and CEPR (gabriella.conti@ucl.ac.uk). Farquharson: Institute for Fiscal Studies (christine_f@ifs.org.uk). Ginja: University of Bergen and IZA (Rita.Ginja@uib.no). Pecher: Department of Health and Social Care (maud.pecher@dhsc.gov.uk).

Acknowledgements: We thank Teresa Steininger for excellent research assistance. We are grateful to Douglas Miller, Lucie Schmidt, Orazio Attanasio, Richard Blundell, Kirill Borusyak, Leon Feinstein, Clement de Chaisemartin, Margaret Leopold, Joanne Pearmain, and Kathy Sylva for their helpful comments. We also received helpful feedback from seminar participants at the University of Bristol, Pompeu Fabra University, UCL Institute of Education, University of Sao Paulo, University of Amsterdam, University of Southern Denmark, DICE, Queen Mary University London, Center for Health Economics at the University Pompeu Fabra and Uppsala University and participants at the Journées Louis-André Gérard-Varet, the International Health Economics Association, Nordic Health Economics Study Group, EEA Annual meetings, Essen Health Economics Workshop, RES Conference, SOLE, AEA Annual Meetings, Economic Demography Workshop, ESPE and the International Workshop on Applied Economics of Education. All remaining errors are our own. The authors gratefully knowledge financial support from the Nuffield Foundation (EYP 42289), the ESRC-funded Centre for the Microeconomic Analysis of Public Policy (ES/T014334/1), the NORFACE DIAL GUODLCCI, and the European Research Council for grants agreement no. 819752 - DEVORHBIOSHIP - ERC-2018COG and ERC-2014-CoG-646917-ROMIA.

1 Introduction

While medical care is a key input in the production of health (Grossman, 1972), more recent theories of health disparities emphasize the role of preventive behaviors (Galama and Van Kippersluis, 2019) and of skills and habits developed in the first years of life (Conti, Heckman and Urzua, 2010; Dalgaard, Hansen and Strulik, 2021). These arguments align with a growing body of evidence showing that early childhood interventions can generate significant and long-lasting benefits for the health of their participants (e.g. Conti, Mason and Poupakis, 2019; D'Onise, McDermott and Lynch, 2010).

Aside from healthcare and parental leave policies, early childhood interventions have followed two main policy approaches to ensure children grow up in a warm, safe and stimulating environment: providing high-quality preschool education outside the home, and/or enriching the family environment (Duncan et al., 2022). Among these early years programs, an interesting type of intervention is 'hybrid' or 'integrated' programs which offer both preschool education *and* family support services, including health services and parenting support. These interventions, which include Head Start, the Chicago Child-Parent Centers, the Perry Preschool Project and the Abecedarian Project, have received considerable attention in the economics literature.

These integrated programs are particularly interesting to advocates of early intervention to promote health for several reasons. First, they aim to promote children's cognitive and socio-emotional development as well as health. As such, they can trigger cross-complementarities in the health production function and be more effective than purely health-focused interventions (Cunha et al., 2006). Second, these programs support the child's environment both inside and outside the home, with the potential to improve several inputs in the health and human capital production function at once. Finally, they have a strong (and largely experimental) evidence base showing that they consistently deliver significant and long-lasting improvements in health.[1]

[1]The most intensive programs, including the Perry Preschool and the Abecedarian projects, have been robustly shown to have large and long-lasting impacts on health (Campbell et al., 2014; Conti, Heckman and Pinto, 2016). Evaluations of Head Start – which operated on a larger scale but is still targeted at disadvantaged children – also tend to find benefits for health in the short and long term. In particular, the Head Start Impact Study RCT finds

This evidence base, however, entirely focuses on programs targeting highly disadvantaged children in the US, where the safety net is significantly weaker than in many other OECD countries. This is important in part because a key channel through which the Perry Preschool and Abecedarian Projects improved the health of their participants is by increasing their take-up of health insurance (Carneiro and Ginja, 2014; Conti, Heckman and Pinto, 2016; Muennig et al., 2009). The existing evidence therefore remains limited in its ability to inform whether similar programs would work if offered to less disadvantaged populations and in contexts with a stronger safety net (including free healthcare). Filling this evidence gap is important to support policy developments both in Europe, where there is a strong tradition of universal programs, and in the US, where free universal healthcare is increasingly debated. It is also particularly timely when integration of children's early years services is a key priority in many OECD countries.[2]

This paper addresses these gaps by evaluating the health impacts of Sure Start, a universal integrated early childhood intervention in England where the National Health Services (NHS) provides free healthcare for all. At its peak, a network of over 3,500 Sure Start centers operated as 'one-stop shops' for families with children under the age of 5, offering health services, parenting support, early education and childcare, and parental employment assistance. Despite being 'one of the most innovative and ambitious Government initiatives of the past two decades',[3] Sure Start has received much less international attention than Head Start, the program it took its inspiration from (Welshman, 2010). This paper is the first rigorous evaluation of this universal program's impact on

short-term benefits for children's cognitive and socio-emotional development, health status, health service use, and health insurance coverage, that fade out by the end of third grade (DHHS, 2010, 2012). Such fade-out 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 centers (Walters, 2015). Other studies have found medium-term benefits for mortality due to conditions that could plausibly be affected by the program (Ludwig and Miller, 2007); obesity (Carneiro and Ginja, 2014; Frisvold and Lumeng, 2011); and depression (Carneiro and Ginja, 2014). In the longer run, Head Start has benefits for adult earnings, education, health insurance coverage, and risky behaviors such as smoking and crime (Anders, Barr and Smith, 2022; Bailey, Sun and Timpe, 2020; Thompson, 2018). Head Start also affected the children of participating parents, in the form of increased educational attainment, reduced teen pregnancy, and reduced criminal engagement (Barr and Gibbs, 2022).

[2]For example, integration of children's services is a key priority for the US government (https://www.dhs.pa.gov/Services/Children/Pages/Integrated-Children's-Services.aspx). In the UK, there is renewed interest in integration of early years services, as reflected in the new 'Family Hubs' initiative building on the legacy of Sure Start.

[3]https://publications.parliament.uk/pa/cm200910/cmselect/cmchilsch/130/130i.pdf

children's outcomes from infancy through adolescence. We ask what impacts access to Sure Start had on children's health, what mechanisms underlie these impacts, and whether it benefited some groups of children more than others.

Our identification strategy leverages the variation across cohorts in the number of centers in the child's Local Authority induced by the program's 11-year rollout across areas of England. Our approach – which controls for small neighborhood fixed effects and cohort fixed effects – is motivated by the fact that the rollout of the program was mostly determined by local deprivation, which is fairly constant over time. In addition to analyzing the determinants of the rollout, we probe the validity of our strategy by testing for pre-trends and checking the robustness of our results to more flexible specifications allowing for differential trends across areas and for heterogeneity in treatment effects across areas and cohorts (Borusyak, Jaravel and Spiess, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021, among others).

Using administrative data covering 21 years of admissions to NHS hospitals (which account for 95% of all admissions), we first estimate the impact of greater access to Sure Start on children's hospitalizations at each age from 1 to 15. We find that greater access to Sure Start *increased* hospitalizations during infancy, but subsequently *reduced* them during childhood and adolescence. At age 1, having access to an extra center per thousand children under 5 increased the probability of a hospitalization by 10% - roughly 6,700 additional hospitalizations a year. The increase in hospitalizations at younger ages is fully compensated by reductions at older ages. At age 5, an additional center per thousand children prevented around 2,900 hospitalizations a year; across all 11- to 15-year-olds, the total was over 13,150 prevented hospitalizations each year.

Changes in hospitalizations may reflect changes in healthcare utilization as well as changes in underlying health, and we perform two additional analyses to tease out whether the program did have an effect on health. Using administrative data, we show that greater access to Sure Start reduced the average length of a hospital stay during middle childhood. Using a nationally representative longitudinal household survey (the UK Household Longitudinal Study), we also show that greater access to Sure Start improved self-reported health and mental health among adolescents.

We interpret both pieces of evidence as strongly suggesting that Sure Start did more than change families' healthcare usage.

A key question with multi-faceted programs such as Sure Start is which service(s) are most effective in driving impacts. No data linking Sure Start services to outcomes are available to directly answer this question, so we bring a number of pieces of evidence to bear on it. First, we argue that different services should lead to different profiles of impacts on hospitalizations for specific causes. Accordingly, we estimate the profile of Sure Start's impacts on hospitalizations for causes that Sure Start may have affected: preventable conditions, infectious illnesses, accidents and injuries, and – among adolescents – mental health. We infer from the results information about which services are likely to be at play. Second, we use another nationally representative longitudinal survey, the Labour Force Survey, to directly test whether the program affected parental employment and family income.

Overall, the findings from these analyses are consistent with Sure Start benefiting children through a number of mechanisms achieved through both health and non-health services. In particular, we find evidence suggesting that Sure Start not only increased parental awareness about children's health and healthcare through information provision, but also strengthened children's immune systems through vaccination campaigns and access to group settings and improved children's behavioral and emotional development by promoting better parenting practices and safer home environments. We rule out services focusing on increasing parental labour supply as an important channel for fostering children's health.

We conclude the paper by testing whether the program had heterogeneous effects across children by gender and deprivation. There are no gender differences until adolescence, when impacts are only visible among boys. Moreover, impacts are strongest among children living in the 30% poorest areas of the country, and entirely null among children living in the 30% richest areas.[4] Tallying these results with descriptive evidence on the socio-economic and gender gradient in the take-up of different services strongly suggests that parenting support services offered by the pro-

[4] Bitler, Hoynes and Domina (2014) also find stronger impacts of Head Start for children at the bottom of the distribution.

gram may have played an important role in driving the improvement in physical and mental health.

These results add to the literature on the health impacts of early childhood interventions. As summarized in Table A.1, to date, the evidence on the health impacts of 'hybrid' programs focuses on the US targeted programs HeadStart, Perry PreSchool and Abecedarian Project (Campbell et al., 2014; Carneiro and Ginja, 2016). The evidence on the health impacts of universal interventions focuses on different types of early interventions, with several papers looking at the long-term impacts of health centers or health visiting programs dating back to the origins of Scandinavian welfare states (Bhalotra, Karlsson and Nilsson, 2017; Bütikofer, Løken and Salvanes, 2019; Hjort, Sølvsten and Wüst, 2017)[5] and several others focusing on policies expanding access to childcare and preschool education (Breivik, Del Bono and Riise, 2021; Hong, Dragan and Glied, 2019; van den Berg and Siflinger, 2022).[6]

The first contribution of our paper to this literature is to show that early childhood programs that combine preschool education and family support services can produce sustained improvements in the health of children even where they already have free access to healthcare. Our analysis of mechanisms strongly suggests that the integration of health and non-health services is key to explaining the hospitalization impacts we find. While we estimate the overall financial benefits of these effects to offset a third of the cost of providing the program, our heterogeneity results suggest that restricting access to these programs to (all families in) deprived areas may be one way to increase their value for money, while avoiding the increase in stigma that could come with programs individually targeted based on family income.

The second contribution of the paper is to provide rare evidence on the profile of impacts of an early childhood intervention through the 'missing middle years', i.e. in between the immediate

[5] Bhalotra, Karlsson and Nilsson (2017) study the introduction of universal post-natal health care, information, and support in the 1930s in Sweden. Bütikofer, Løken and Salvanes (2019) evaluate the very long-run impacts of a 1930s program of mother and child health centers and post-natal home visiting in Norway. Hjort, Sølvsten and Wüst (2017) study the long-term health impacts of a universal health visiting intervention in Denmark for all infants.

[6] van den Berg and Siflinger (2022) finding that subsidizing childcare in one region of Sweden at age 1 decreases the number of medical visits at ages 4-5 and 6-7. Breivik, Del Bono and Riise (2021) find that increased access to universal childcare during the 1970s in Norway increased the used of primary and specialist health care services for women at ages 30 to 47, but not for men. Hong, Dragan and Glied (2019) find that the New York City's universal pre-kindergarten program increased the probability that a child is diagnosed with asthma or with vision problems, receives treatment for hearing or vision problems, or receives an immunization or screening during the pre-kindergarten year.

impacts and the longer-term effects that most existing studies estimate (Almond, Currie and Duque, 2018). This evidence is important not only for the cost-benefit analysis of the program, but also because the persistence of the effects in the post-eligibility years provides a stronger basis for predicting longer-term impacts - a key concern of policymakers seeking to justify spending on early intervention. Moreover, as we illustrate here, tracing out the profiles of program impacts through the medium term can shed crucial light on the mechanisms through which multifaceted programs like Sure Start work. Through this analysis and our heterogeneity analysis, the paper contributes to understanding *why* integrated programs work and *for whom*, an area Duncan et al. (2022) highlight as needing much greater research in their recent review.

The third contribution of the paper is to study a major program that has received relatively little attention so far. While there are two government-commissioned studies of Sure Start, neither of them use methodologies that support a robust causal interpretation of the links they find between children's outcomes and access to and usage of Sure Start.[7] Our paper proposes a robust quasi-experimental design using administrative data, which allows us to examine the causal impacts of the program much beyond the time horizon considered by previous evaluations.

2 The rollout of Sure Start

First introduced in 1999, Sure Start was conceived as an area-based intervention whose services would be available to all families with a child under five in the neighborhood of the center (without individual means-testing). Sure Start was given a budget of £450 million over the period 1999-

[7] The first, the National Evaluation of Sure Start (NESS), collected data on children living in neighborhoods served by the earliest Sure Start Local Programmes in 2001 and compared these children to others surveyed in an earlier national survey (conducted in 2000) who lived in areas not served by the program. The NESS found an increase in parent-reported hospitalizations at 9 months, an increased prevalence of immunizations and a reduced probability of accidental injuries 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 neighborhoods (NESS, 2005, 2008, 2010). At age 5 and 7, it also found that this group had better family functioning (e.g. better home learning environment, less chaotic homes), greater social skills, and lower behavioral issues than the group not living in a neighborhood with a Sure Start center. The second study, the Evaluation of Children's Centres in England (ECCE), was run in 2011 and collected data on a sub-sample of Children's Centers and their users, and compared the outcomes of children whose families chose to use the services with varying frequency. The ECCE study found no significant association between the extensive margin of using Sure Start services and child health, but did conclude that more intensive service use was associated with fewer externalizing behavior problems, higher child physical and maternal mental health, and improved family functioning (ECCE, 2015).

2002 (roughly £700 million, or $800 million, in 2021 prices). This funding was to initially rollout 250 Local Programmes (SSLPs) in highly disadvantaged areas (Melhuish et al., 2008; Pugh and Duffy, 2010). To decide which areas would get funding to open a Sure Start center, a national Sure Start Unit (SSU) developed a set of guidelines for the rollout. The initial 60 'trailblazer' districts invited to submit an application were selected based on three characteristics: the local level of deprivation, the incidence of low birth weight and of teen pregnancy. By November of that year, 15 SSLPs had opened their doors (DfEE, 1999). The program proved so popular that it was quickly expanded. A year after it began, the target number of SSLPs was doubled from 250 to 500 (Eisenstadt, 2011).

Four years later, the government announced that Sure Start would transition from a program for disadvantaged neighborhoods into a universalized offer, with "a children's center in every community" by 2010 (DfES, 2003). This expansion also included a rebranding (from 'Local Programmes' to 'Children's Centres') and a greater role for central government in setting out a 'Core Offer' of services (Lewis, 2011), which we describe below. In current prices, the real-terms annual budget for Sure Start rose from about £520 million (about USD 600 million) to £1.8 billion (USD 2.1 billion) at its peak in 2009-10, or about a third of overall spending on programs for the under-5s in England, attesting of the national importance of the policy (Britton, Farquharson and Sibieta, 2019).

The Department for Education and Skills (DfES) had overall responsibility to establish 3,500 children's centers by 2010. The rollout was intended to be driven entirely by local deprivation, with three well defined phases (House of Commons, 2010). Between 2004 and 2006, there would be approximately 800 Phase 1 centers to offer full coverage of the 20% most disadvantaged neighborhoods. Of these, around 500 would grow out of existing SSLPs while the rest would be new centers. In Phase 2, between 2006 and 2008, 1,700 new centers would open in the 30% most disadvantaged neighborhoods. Finally, Phase 3 of the rollout would complete the universalization of the program, through the opening of another 1,000 centers in the remaining 70% of areas.[8]

[8]While the national government retained overall control of this phased approach, Local Authorities (LAs) also gained more control over decisions about where specifically centers within their jurisdiction were located. LAs were

Following the rollout plans, the period between 2005 and 2010 saw a rapid increase in the number of Sure Start centers, with the rollout described by many as 'haphazard' and 'too fast' (Morris, Barnes and Mason, 2009; Prowse, 2008). By 2010, the overall number of centers reached 3,500, with each center serving a local population of between 600 and 1,200 children depending on the location and level of need (see Appendix Figure A.1 for the full trend in number of centers). The maps in Appendix Figure A.2, which show the location of Sure Start centers in 2000, 2004, 2006, and 2008, suggest that the deprivation-based guidelines for each of the phases were taken seriously.

Sure Start since 2010 Elections in 2010 brought a new government, which de-prioritized Sure Start. Between 2011 and 2019, government spending on Sure Start fell by over 60% (Britton, Farquharson and Sibieta, 2019). The core mission of the program became less prescriptive, allowing each center to focus on the outcomes it wanted to achieve for young children and families (Smith, 2018). Following the removal of the funding ring-fence in April 2011, local authorities responded to these cuts in different ways, with some subsidizing Sure Start services from other budget lines while others consolidated several centers into one, cut back on services or shut down centers (Smith, 2018). In light of these important changes, this paper focuses on the 1999-2010 period during which the program expanded and delivered a more consistent service offer.[9]

3 Sure Start services and expected effects

3.1 Which services did Sure Start offer?

The overarching aim of the Sure Start initiative was to improve outcomes for young children. Its approach to child development was based on the recognition that child development is multi-

allocated targets and funded to deliver centers based on their under-five population and their level of deprivation.

[9] Our empirical strategy, presented in subsection 4.2, exploits the variation in access to Sure Start resulting from the rollout to identify the effect of increased access to Sure Start on children's hospitalizations. In principle, the reduction in Sure Start access resulting from center closures could also be used within such an empirical strategy. However, we refrain from doing so because, given the freedom with which LAs could respond to funding cuts, center closures are more likely to be endogenous than their openings. Moreover, we only have imperfect information about center closures, which does not capture 'hollowing out' of services in centers that technically remained open.

dimensional and that the needs of families, particularly disadvantaged families, often span many traditional areas of support. As a result, Sure Start offered a *range* of services to support children and their parents *integrated* within each Local Programme or Children's Centre.

While there is no centralized record of the services each center offered, several sources point to the fact that both Sure Start Local Programmes and Children's Centres offered services across four main domains: health, parenting support, childcare, and parental job assistance. Sure Start Local Programmes had more latitude in choosing the services they would offer to meet local need than Sure Start Children's Centres did.[10] From 2004 Children's Centres in the 30% most deprived areas (Phase 1 and Phase 2 SSCCs) were required to meet a 'Core Offer'. This consisted of outreach to parents; early education and childcare; family and parenting support; child and family health services (such as antenatal support); and links with JobcentrePlus, a government-funded employment agency and social security office that aims to help people of working age find employment (Lewis, 2011). For centers serving the 70% least disadvantaged areas ('Phase 3' SSCCs), centers could meet some of the requirements of the Core Offer by developing referral links to other services; still, all centers were expected to offer activities for children, health and outreach services, and links to Jobcentre Plus.

3.2 How did Sure Start change early years services provision?

Sure Start changed the counterfactual provision of early years for families with children under five in three main ways. First, it brought together existing services under a single roof and streamlined referrals to more specialized services, such as services for children with Special Education Needs and Jobcentre Plus. The fact that parents could access a range of services during one visit may have increased the take-up of existing services as a result of this integration.

[10] A survey conducted in 2003-2005 found that the largest single area of spending was play, learning and childcare, which accounted for a third of the SSLPs' budget (Meadows, 2011). A fifth of spending was aimed at community healthcare; this funding was used to supplement existing health services or to provide services that were not available through the public healthcare system (e.g. services for postnatal depression). Another fifth of the budget was spent on parenting support, and a sixth of spending went on outreach and home visiting. Support for children with special needs accounted for 7% of spending, and the remainder of the budget was spent on premises costs and other activities (Meadows, 2011).

Second, Sure Start provided new programs to meet local unmet demand (DfEE, 1999). Among those, Sure Start provided health services, such as breastfeeding support and support for postnatal depression, which would not have been available through the public healthcare system. It also provided parenting support, such as evidence-based parenting classes and drop-in play session, for free or a very low price.

Third, Sure Start may have changed families' childcare use. Sure Start was rolled out at the same time as another policy called the 'Free Entitlement', which offered every 3- and 4-year-old up to 15 hours a week of fully subsidized early education. This meant that much of the funding for childcare hours in practice came through the (separate) free entitlement budget. Sure Start centers were involved in signposting to the program, and some centers also chose to deliver some funded childcare places on site.

While there is no data available on the number of places created in SSLPs or SSCCs, we would not expect the opening of Sure Start to have significantly increased the total supply of childcare places in England. By the early 2000s, already over 80% of 3- and 4-year-olds were taking up some form of center-based childcare, indicating that the supply of places was already high. By 2007, this proportion reached almost 100%, but most of this increase was supported by the expansion of private nurseries at the time rather than by the expansion of Sure Start centers (Blanden et al., 2016). Estimates from a survey conducted in the early 2010s (described further below) suggest that around 6% of families took up formal childcare at a Sure Start center; a further 4% used childcare settings that they had been signposted to by someone at their Sure Start center. Among 3- and 4-year-olds, 8% used childcare provided by Sure Start and a further 7% had been signposted.

However, the rollout of Sure Start could have had an effect on the type of nurseries children attended. Phase 1 and Phase 2 Children's Centres were required to deliver some childcare on site, with more stringent requirements on staff qualifications than private nurseries.[11] Sure Start may therefore have led families to substitute away from private nurseries into Sure Start nurseries, where they might have been able to access childcare of higher quality than they would have otherwise.

[11] Phase 3 Centres were not required to provide early learning and childcare places but could do so if the need arose (House of Commons, 2010).

3.3 Descriptive evidence on the take-up of Sure Start services

In absence of individual-level information on the take-up of Sure Start throughout the rollout period, we use two surveys to provide descriptive evidence of families' use of the program at two time points: very early in the rollout and at its end. First, we use the Millennium Cohort Study, a nationally representative cohort study of children born in 2000-02 in the UK. In its first wave, administered when children were 9 months old, parents were asked about their awareness and usage of Sure Start. 28% of families report knowing of Sure Start and 5% of them report having used Sure Start. As we show in Table A.2, these national figures mask a lot of variation across families and areas. In particular, Sure Start awareness and usage are significantly higher in Local Authorities with a greater concentration of Sure Start centers. Controlling for a host of child and family characteristics, we estimate that one extra SSLP per 1000 children age 0-4 is associated with a 13 percentage point increase in the probability of using Sure Start.

The second data source with information on take-up is a survey data collected as part of one of the government-commissioned evaluation of Sure Start, the Evaluation of Children's Centres in England (ECCE). The data is a longitudinal panel starting in 2011 following parents who had an infant registered with one of the 128 Phase 1 and Phase 2 Children's Centres selected to participate in the study.[12] Because by then almost all families were automatically registered with a center from the birth of their child, this sample can be used to analyze the take-up of Sure Start services in centers located in the 30% most deprived areas of the country (where, by then, about 50% of centers were located).[13] Based on these data, Figure 1 shows that, across all ages, the services most likely to be used were health services and parent-child services (such as drop-in play sessions or baby

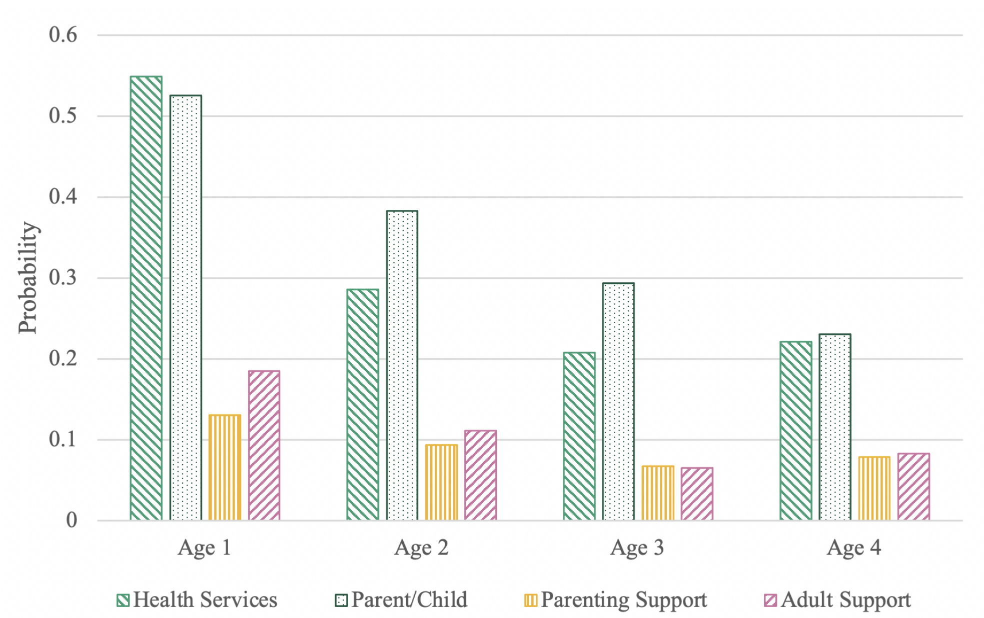
[12]Data was collected at three points in time. In the baseline survey, 5,717 parents were surveyed through face-to-face interviews when their child was aged 9-18 months old. All parents who had agreed to be re-contacted and provided a telephone number were invited to take part in a second survey, and 3,588 phone interviews were contacted in 2013 when the selected child was aged about 2 years old. Another round of data was collected through face-to-face interviews of 2,692 parents whose selected child was now about 3 years old. The analysis reported in the paper pools all years of data together so as to reflect almost the entire range of eligible ages. See Goff and Chu (2013) for further information.

[13]By this time, there was a well-functioning system of referrals in place to ensure that a very high proportion of families registered with their local Children's Centre, and centers were judged by the regulator on the proportion of their local population who they 'reached'. However, registration is distinct from accessing services: around a quarter of this population did not access any services.

classes), though their take-up rapidly declined with the child's age. By contrast, the proportions of families using parenting support and adult support was lower, but more stable across age groups.

As we show in Appendix Table A.3, different services also attracted different types of families. At young ages, families with higher socio-economic status were more likely to use Sure Start services, but among two- and three-year-olds take-up was higher among lower-SES families. The exception to this is adult support services, which more intensively used by the low-SES regardless of child age. We see few gender differences in take-up, though the parents of girls are less likely to use Sure Start parenting services. We return to this when we discuss the program's impact heterogeneity by gender and deprivation in section 7.

Figure 1: Take-up of services in Phase 1 and Phase 2 centres



Notes: The figure shows the probability that families in the ECCE sample report making use of each of the four main types of services, by the age of the focus child. This is based on pooling all three waves of ECCE data. Health services include e.g ante-natal classes, breastfeeding groups, midwife/health visitor drop-in session or clinic; Parent/child services include e.g. stay and play, or play and learn drop-in sessions, organized sport or exercise for babies or children, toy libraries: Parenting support include peer support groups (parents supporting other parents), parenting classes, and specialist family or parenting support ; and Adult support includes e.g. benefits and tax credits advice, housing or debt advice, employment support, Basic IT or jobs skills course.

3.4 Expected effects of Sure Start on health and hospitalizations

Given the variety of services offered by Sure Start, there are numerous channels through which the program could have affected children, and these impacts could have changed over time. In this section, we discuss how each of the four main groups of services could have affected health and hospitalizations from infancy through adolescence.

Health services Sure Start centers offered a range of health services, including ante-natal and post-natal support for mothers and babies; advice on accident and injury prevention; advice on obesity, diet and nutrition; and support for mental health and for families with disabilities (e.g. DfE, 2010; DfES, 2003). Sure Start did not substitute for primary care provision (which is freely delivered in England by General Practitioners), but rather enhanced access to health support and information.

These supplemental health services may have affected hospitalizations through two main channels. The first is screening children for conditions and referring families to appropriate healthcare, leading to an increase in hospitalizations for preventable and manageable conditions[14] in the short term (early years) and a decrease in hospitalizations for the same conditions in the longer term. The second mechanism is enhancing health-promoting parental behavior and the safety of the home environment, leading to a reduction in hospitalizations at all ages. Since Sure Start provided information about accident prevention, 'child-proofing' and safety in the home, we could expect this reduction to be particularly important for accidents and poisonings. Given the nature of the advice focused on very young children, effects may be more pronounced during the early years, although if this information led to sustained changes in parenting behavior, those effects may be longer-lasting.[15]

[14]Specifically, we consider Ambulatory Care Sensitive conditions. This group includes conditions that are typically managed with primary care (such as asthma), conditions that are largely preventable (e.g. gangrene) and vaccinable conditions (such as measles).

[15]Information was also provided about diet and nutrition, and we could also expect this advice to decrease the incidence of obesity. In a preliminary version of this paper (Cattan et al., 2019), we test for this mechanism directly by employing a similar research design and administrative data on weight and height of all children in primary school at ages 5, when there is an census of height and weight through the National Child Measurement Programme. We found

Parenting support and parent-child services Centers provided evidence-based parenting classes (such as the Triple P and Incredible Years programs) to improve family functioning and positive parenting skills, often with a particular focus on children's mental health and emotional and behavioral issues. They also provided a range of parent-child activities, such as fathers' groups, baby massage and drop-in play sessions, aimed at strengthening parent-child relationships.

These services did not target (physical) health specifically, but may nevertheless have had indirect benefits by activating cross-productivities between behavioral and emotional development and health (Cunha et al., 2006). By strengthening child-parent attachment and parenting practices, these services may have led to healthier emotional and behavioral development (Case and Paxson, 2002). Calmer and less fidgety children have been shown to be less prone to injure themselves and may be easier for parents to care for (Hoare and Beattie, 2003). Early intervention to improve parent-child bonds may also reduce the chances of parental neglect and maltreatment (Avellar and Supplee, 2013; Eckenrode et al., 2017). Later in life, stronger emotional and behavioral regulation could help children's mental health and reduce their exposure to accidents and injuries from risky or aggressive behavior.

As a result, we expect these services to reduce hospitalizations for accidents and injuries during the early years. To the extent that early benefits for parenting or child development persist, these reductions in hospitalizations could be lasting. Moreover, if Sure Start improved children's emotional development, we would also expect a reduction in hospitalizations for mental health-related causes. However, since the prevalence of mental health-related hospitalization is essentially zero before adolescence, we would not expect to detect an effect on mental health hospitalizations during the early years or childhood in the data we have available.

Childcare and group-based sessions If Sure Start facilitated the take-up of high-quality childcare, it could have affected health in two ways. First, high-quality childcare can benefit emotional and behavioral development (Heckman, Pinto and Savelyev, 2013). As such, we could expect a reduction in hospitalizations for accidents and injuries similar to those resulting from parenting

no evidence of effects of the program on obesity. We rule out this mechanism going forward.

support and parent-child services. Second, childcare (and, to a lesser extent, other group-based activities) increased the time children spent around other children and hence their potential exposure to infectious diseases. In the short run, this might have led to an increase in the number of sickness episodes. But early exposure to a variety of pathogens also helps to build up the immune system, which might have benefits in the longer run (Henderson et al., 1979; van den Berg and Siflinger, 2022). In this case, we would expect hospitalizations for infections to increase in the short term and drop in the medium term. We would not expect these negative effects to be particularly long lasting, however; all children go to school from the age of 5 in England, meaning that children who were not exposed to Sure Start should see their immune systems catch up once they start spending more time with others.

Adult support The last major set of services offered by Sure Start aimed to support parents, especially in their effort to gain employment. A subsequent increase in parental employment could affect children's health by allowing parents to buy more and/or higher quality inputs, such as more nutritious food (Carneiro and Ginja, 2016). However, employed parents may have less time to spend on health-improving activities (e.g. cooking a home-made meal or accompanying children to the doctor). Parents shifting into employment could result in children spending more time in childcare. Since these channels push in different directions, the overall effect of Sure Start's employment services on children's health and hospitalizations is ambiguous. This channel does not lead to clear testable predictions in the hospitalization data, therefore we present a separate estimation of the effect of Sure Start on parental employment using data from the Labour Force Survey (LFS) (Appendix F).

Appendix Table A.4 summarizes this discussion by describing the potential impact of each type of services on hospitalizations in infancy, middle childhood and adolescence. For clarity, we have highlighted effects on the same type of hospitalizations in the same color. As the Table makes clear, the direction of Sure Start's impact on hospital admissions is expected to differ based both on the cause of hospitalization and the age of the child. This means that the *overall* impact on hospitalizations is expected to be ambiguous, particularly earlier in life when higher admissions for infections or preventable conditions may offset reductions in admissions for accidents and injuries. As children grow older, however, we would expect the effect of Sure Start on overall admissions to be more clearly negative. Because we do not have data on service take-up to link to hospitalization data, we are not able to probe the mechanisms through which Sure Start worked directly. Instead, guided by the discussion above, after presenting estimates of impacts on overall admissions, we present estimates of impacts on cause-specific hospitalizations at different ages to infer information about what mechanisms were most likely at play (section 6).

4 Data and empirical strategy

4.1 Data sources

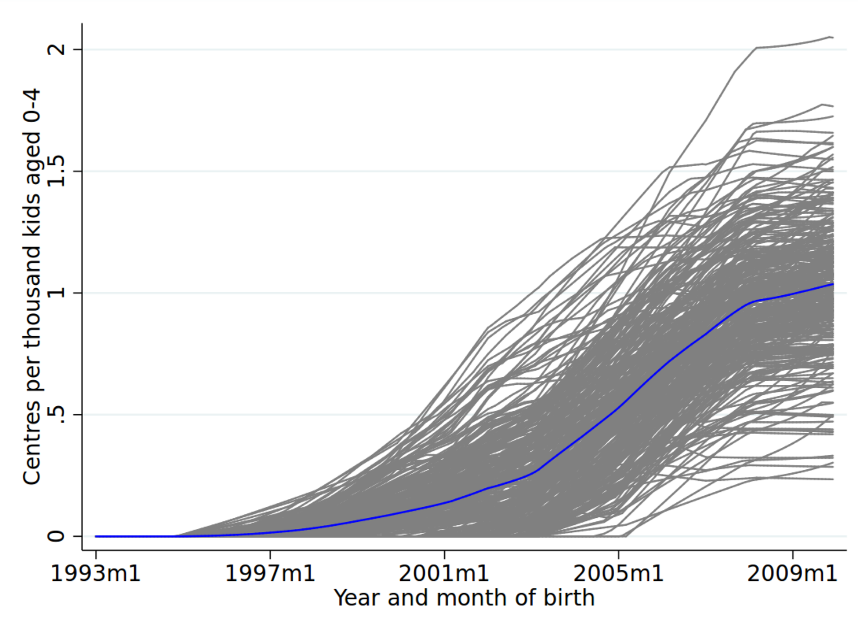
Data on Sure Start facilities To measure our treatment variable, we use a unique dataset containing the exact address and date of opening of each Sure Start Local Programme and Children's Centre between 1999 and 2010. Based on this information, we construct our measure of access to Sure Start SS_{dq} that varies across Local Authority d and quarter of birth q (our cohort dimension).[16] Specifically, we define SS_{dq} as the average number of centers per thousand children aged 0-4 that were open during the first 60 months of life of a child born in quarter q and living in Local Authority (LA) d.[17] The treatment variable SS_{dq} is the number of centers *per* 1,000 children of eligible age in order to account for the fact that some Local Authorities are larger than others and each center was designed to reach about 800 children (House of Commons, 2010). When estimating models with an outcome measured before age 5, we define SS_{dq} as the average number of centers per thousand children aged 0-4 that were open between the child's birth and the age at which the outcome is measured. Figure 2 plots this variable for each of the 323 LAs in England

[16] By 'quarter' of birth, we refer to the combination of a year and quarter. Given that our maximum sample includes children born from January 1 1993 to December 31 2006, we have children born in 52 different quarters of birth or cohorts in the data.

[17] There are 326 Local Authorities (LAs) in England. We exclude three of them from the analysis (the Isles of Sicilly, City of London, and West Somerset), which are very small areas with few children aged 0-4 and appear as outliers in terms of Sure Start coverage.

(in gray) and superimposes its average (in blue) across LAs. Across the cohorts we consider, the number of SS centers per thousand children aged 0-4 increased from 0 to an average of just over 1.[18]

Figure 2: Average coverage over the first 60 months of life, by local authority and month and year of birth



Notes: Each gray line represents one of 323 Local Authority (LA) districts in England (excluding the Isles of Scilly, City of London, and West Somerset). The blue line shows the average for all of England. The lines plot the average Sure Start coverage (centers per thousand children aged 0-4 in the LA) over the first five years of life for children based on their month and year of birth. Source: Authors' calculations using data from the Department for Education and ONS population estimates.

Data on hospitalizations A large part of our analysis focuses on children's hospitalizations, which we measure using the Hospital Episode Statistics (HES), an administrative dataset tracking the universe of patients using English public hospitals. Data on inpatient admissions has been collected since April 1997 and we have data up to March 2018. The inpatient data tracks all hospital admissions, providing information on the admission, discharge and clinical diagnoses (up to 20

[18] Our treatment variable does not distinguish between Local Programmes and Children's Centres. While the opening dates of all centers are precisely known, pooling in this way requires us to make an assumption when SSLPs transitioned into SSCCs, which has not been recorded in the data. Since over 90% of SSLPs had transitioned into Children's Centres by 2006 (NAO, 2006), 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 in December 2006.

for each patient).[19] It also include demographic information on the patient's sex, ethnicity, date of birth, and the Lower-level Super Output Area (LSOA) of residence at the time of admission.[20] To maximize comparability across cohorts, we restrict our sample to children born within 5 years of the initial announcement of Sure Start (i.e. those born in 1993 or later) and to children who could only have been exposed to Sure Start before the 2010 change in policy (i.e. those born in 2006 or before).

To create our outcomes of interest, we include one record per hospitalization (though patients may have several 'episodes' under different physicians during a single spell of hospitalization) and exclude admissions related to the birth of a child. We then construct counts of all-cause and (primary) cause-specific admissions for each neighborhood (defined at the LSOA level), quarter of birth, sex and age of admission. Cells without admissions are assigned zero. Because a large fraction of cells have zero admissions, we define our main outcome of interest as $D^{ya}_{sql(d)}$, an indicator for whether there is any hospitalization of type y at age a for children of sex s born in quarter q and residing in neighborhood l (of LA d).

Other data sources We use a number of additional data sources to estimate the impact of Sure Start on a wider set of outcomes. Specifically, we use the UK Household Longitudinal Study (UKHLS) and the Labour Force Survey (LFS), two nationally representative household panel surveys, to estimate the impact of greater access to Sure Start on self-reported health and mental health and on parental employment, respectively. Finally, and as mentioned before, we make use of two additional surveys, the Millennium Cohort Study (MCS) and the ECCE survey, to provide descriptive evidence on families' awareness and take-up of Sure Start services. We provide further details about each of these data sources when we describe the relevant results throughout the paper.

[19]In this context, inpatient admissions include day cases who are admitted to a hospital bed as well as those who stay overnight. There is a separate register for emergency room attendance, but these data are only considered reliable from April 2007, so there is less scope to look at the impacts of Sure Start across the entire life-cycle of the program. Similarly, the register for outpatient data is only reliable from April 2006.

[20]The LSOA is a very small geographic unit. There are around 33,000 LSOAs in England, and the average LSOA has a population of around 1,500 residents. LSOAs are a unit of statistical rather than practical geography, so there are no administrative or electoral responsibilities that are conducted at the LSOA level. However, LSOAs nest within Local Authorities (LA), which do have a role in administering a wide range of policies.

4.2 Econometric specification

Our aim is to estimate the effect of increased access to Sure Start on children's hospitalizations (for any cause and for specific causes). To do so, we exploit the variation in potential exposure to Sure Start across birth cohorts and Local Authority generated by the Sure Start rollout and displayed in Figure 2 in a standard difference-in-differences framework. We operationalize it by way of a two-way fixed effect model, where we control for: i) birth cohort fixed effects to account for secular trends in hospitalization, and ii) neighborhood fixed effects to account for systematic differences in time-invariant area characteristics, such as deprivation, which we know are correlated with both the Sure Start rollout and hospitalizations.

Our main estimating equation has the following specification:

$$D^{ya}_{sql(d)} = \delta^{ya} SS_{dq} + \beta^{ya} X_s + \alpha^{ya} Pop_{al} + \gamma^{ya}_q + \pi^{ya}_{l(d)} + \upsilon^{ya}_{sql(d)}, a = 1, ..., 15 \quad (1)$$

where $D_{sql(d)}$ and SS_{dq} are defined as above, X_s is a female dummy. We control for Pop_{al}, the number of children of age a in neighborhood l, because the probability of observing non-zero hospital admissions rises with population size. γ^{ya}_q is a set of cohort of birth fixed effects defined at the year-quarter level. The model includes a set of over 32,000 neighborhood fixed effects $\pi^{ya}_{l(d)}$, which account for time-invariant unobserved heterogeneity across areas. Finally, the error term is denoted $\upsilon^{ya}_{sql(d)}$. Since the differences in population across neighborhoods are relatively modest, our main results are not weighted; however, we obtain similar results when weighting each cell by population size (see Table A.5).

In section 5, we report the estimates of parameters in equation (1) for admissions for any cause and ages $a = 1, ..., 15$. In subsequent sections, we re-estimate this model for cause-specific hospitalizations and for different subgroups. The parameter δ^{ya} is an Intention-To-Treat (ITT) parameter, as it measures the effect of increasing access to rather than actual use of Sure Start. Since Sure Start was designed as an area-based intervention, this parameter also corresponds to the relevant parameter to compute the net benefits of the policy.

For all models considered, we present robust standard errors clustered at the level of Local Authority (LA) at the time of admission to account for autocorrelation in the outcomes (Bertrand, Duflo and Mullainathan, 2004). We study a relatively large number of outcomes so we also report the results of a stepwise multiple hypothesis testing procedure that controls for family wise error rate (Romano and Wolf, 2005).[21] In line with our discussion of expected effects, when applying this correction we consider the different phases of child development and test simultaneously the impacts for three age groups: 0 to 4 (early years), 5 to 10 (middle childhood) and 11 to 15 (adolescence).

4.3 Validity of empirical strategy

The interpretation of the parameter δ^{ya} as the causal effect of increasing *access* to Sure Start relies on three assumptions: (1) that greater access to Sure Start increases the probability of participation; (2) the "parallel trends" assumption that the rollout of Sure Start across LA was uncorrelated with time-varying unobservable determinants or shocks to hospitalizations (captured in $\upsilon^{ya}_{sql(d)}$); and (3) that families did not locate selectively to be closer to Sure Start centers as they were rolled out. Our discussion of the take-up of Sure Start services in subsection 3.3 provides descriptive evidence that greater access to Sure Start was related to greater use of its services, thus validating (1). The rest of this subsection therefore concentrates on providing evidence supporting the validity of assumptions (2) and (3).

4.3.1 Parallel trends

The parallel trends assumption requires that the rollout of Sure Start was uncorrelated with unobserved time-varying determinants of hospitalizations. As discussed in section 2, official guidelines about the Sure Start rollout indicate that new SSLPs were prioritized in areas with relatively high deprivation, high teenage pregnancy rates, and a high proportion of low birthweight births (under

[21]Specifically, we use the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005) to account for testing several hypotheses simultaneously; this is an iterative rejection/acceptance method for a fixed level of significance. We use 500 block-bootstrap replications to obtain the adjusted critical values (the block is the LA).

2.5kg), while the rollout of SSCCs (from 2004 onwards) was mostly determined by area deprivation. Nevertheless, policymakers' decisions over where and when to open new centers could have been influenced by other factors affecting the supply and demand for the centers. Our identification strategy would be under threat if we found that these factors varied differently across areas and also affected children's hospitalizations. We present several checks to assuage concerns about such threat.

Determinants of the rollout First, we analyze the determinants of the rollout. In addition to data on known determinants of the rollout (overall deprivation, teenage pregnancy rates, and proportion of low birthweight births), we consider a range of other factors that might have influenced where and when centers opened. We gathered data on LA and year-level information about local labor market conditions (male and female weekly full-time earnings and the claimant rate for unemployment benefits); potential demand (number of children aged 0-4 in the LA and proportion of children aged 0-4 looked after); health indicators (infant mortality rate and number of General Practitioners per 1000 inhabitants in each LA); local services (proportion of 3-year-olds taking up a funded part-time childcare place, potentially but not necessarily in Sure Start centers); and political variables (share of local council seats held by the Labour party, which was the party in power nationally during the expansion of this policy). Importantly, while all these factors could have affected decisions about where to open new centers, they could all potentially affect hospitalizations either directly or indirectly.

We start by showing the extent to which the change in Sure Start coverage is explained by time-invariant characteristics and time-varying demand and supply factors. We regress the Sure Start coverage rate (defined at the LA and quarter of birth cohort level) for potentially exposed cohorts (born between 1996 and 2006) on LA fixed effects, cohort fixed effects (i.e. year-quarter fixed effects), and the potential determinants of the SS rollout described above. Area and cohort fixed effects explain 86% of the variation in SS coverage, while only 4% of the variation can be explained by time-varying demand and supply factors. This suggests that the rollout was mostly determined by time-invariant area characteristics, which our empirical strategy controls for via neighborhood fixed effects.

Following Bhuller et al. (2013), we further investigate how the year-to-year, within LA variation in SS exposure at the core of our empirical strategy correlates with baseline area characteristics by estimating the following equation:

$$\Delta SS_{dq} = \rho_d + [\beta_q \times c_{d,1998}]' \Phi_q + \epsilon_{dq} \tag{2}$$

where $\Delta SS_{dq} = SS_{dq} - SS_{dq-1}$ and $c_{d,1998}$ is the vector of LA characteristics described above and measured in 1998, the year preceding the opening of the first SSLP. We plot the estimated coefficients Φ_q and their 95% confidence intervals in a series of graphs shown in Appendix Figure A.3. Consistent with official guidelines about the SS rollout, we find that the expansion of Sure Start coverage is positively associated with deprivation, teenage pregnancy rate, and proportion of low birth weight births for cohorts born between 1996 and 2002 (i.e. cohorts mostly exposed to SSLPs). We also find a positive and significant correlation between the expansion in SS coverage and local unemployment rate until 2003 (conditional on deprivation level), which is perhaps unsurprising given that one of the core objectives of Sure Start was to help increase parental employment. From 2003 onward, there is no correlation between any of the variables we consider and the local rollout of the program.

Pre-trends We complement this analysis of the rollout by testing whether the timing and intensity of the rollout was systematically correlated with hospitalizations in the pre-intervention period. We use hospitalization data for cohorts who could not have been treated by the time we see them admitted in the HES. Because our hospitalization data starts in 1997 and the first Sure Start Local Programme opened late in 1999, the number of cohorts we can use to estimate pre-trends varies with the age of hospitalizations we focus on.[22] To keep the sample used to test for pre-trends

[22]For example, for admission at 1, we can use hospitalizations of one-year-olds born in 1996, 1997 and 1998 (and observed in year 1997, 1998 and 1999 respectively) to test the existence of pre-trends in age 1 admissions. For admission at later ages, we have more cohorts available. For example, for age 7 admission, we can use hospitalizations

as consistent as possible across admission ages, we use all available untreated cohorts born after 1990.

We study whether pre-trends in hospitalizations are correlated with (i) the timing of the rollout (i.e. when the first center opened) and (ii) the intensity of the rollout (i.e. the coverage) by estimating the following models on the sample of untreated area-cohorts cells for each age $a = 1, \ldots, 15$:

$$D^{ya}_{sql(d)} = \sum_{l=1}^{l=L} \delta^{ya}_l \mathbf{1}[SS_{d,q+l} > 0] + \beta^{ya} X_s + \alpha^{ya} Pop_{al} + \gamma^{ya}_q + \pi^{ya}_{l(d)} + \upsilon^{ya}_{sql(d)} \quad (3)$$

$$D^{ya}_{sql(d)} = \sum_{l=1}^{l=L} \delta^{ya}_l SS_{d,q+l} + \beta^{ya} X_s + \alpha^{ya} Pop_{al} + \gamma^{ya}_q + \pi^{ya}_{l(d)} + \upsilon^{ya}_{sql(d)} \quad (4)$$

where L is the maximum number of pre-trend coefficients we estimate in the regression. Because we run these regressions on the sample of cohort-area observations that have not been treated yet, the reference category is cohort L-1. The estimates of these regressions are reported in the series of graphs in Figure A.4 and Figure A.5, which show that there was no systematic relation between the timing and intensity of the rollout and hospitalizations in the pre-treatment period.

Robustness checks to differential trends across areas Overall, the evidence presented above suggests that the variation we use to estimate the impact of increased access to Sure Start on hospitalizations is unlikely to be correlated with unobservable determinants of hospitalizations. Nevertheless, in subsection 5.2, we present the results of a number of additional specifications to provide further reassurance that differential trends across areas are not driving our results. First, we estimate the specification including LA-specific linear trends, which we estimate on the pre-treatment period. Moreover, we also estimate a specification that includes interactions between a cohort trend and the three official determinants of the rollout (deprivation level, teen conception rate, and incidence off low birth weight all measured in 1998), which we showed affected the rollout of Sure Start. As we show below, our estimates are very robust to the inclusion of these flexible trends in the model, which provides further reassurance that the parallel trends assumption

of seven-year-olds born in 1990 to 1995.

is likely to hold in our setting.

Robustness to confounding policy shocks A final threat to the validity of the identifying assumption may arise if hospitalizations were subject to other, confounding policy shocks that correlated with the rollout of Sure Start. One crucial set of potential confounders is the local service offer. Over the decade that Sure Start was rolled out, the national government also made a number of reforms to the benefit system, the health system and the early years system. In-work benefits became more generous while out-of-work benefits were reduced (Gregg, 2008); health spending rose from 5% to 7.5% of GDP (Stoye and Zaranko, 2019); and the government introduced the 'Free entitlement' to a free part-time childcare entitlement for 3- and 4-year-olds (Blanden et al., 2016).

While our empirical strategy allows for cohort effects, many of these reforms may have impacted local authorities differently. We will therefore present a robustness check that controls for a range of time-varying LA-level characteristics, including the rollout of funded childcare places; the number of General Practitioners per capita, a proxy for health service availability; and local labor market characteristics (to reflect changes in the benefit system incentivizing employment). In addition to these measures of the policy environment, we will incorporate a wide range of other characteristics that may be related to both the rollout of Sure Start and the incidence of hospitalizations. These include local demographics; vital statistics; and labor market characteristics (see Appendix B for the full list of variables and sources). If any of these characteristics is confounding our results - or is correlated with another unobserved characteristic that is confounding our results - we would find that our results are not robust to these specifications. We present evidence below that this is not the case, thus providing reassurance that the effect of Sure Start access we estimate is not confounded by the effect of other policies (see subsection 5.2).

4.3.2 Selective migration

A final requirement for the validity of our empirical strategy is that there is no selective migration of families into high-coverage areas. Using data from the British Household Panel Survey, we show in Appendix C that there is no relationship between migration and Sure Start coverage. Furthermore, the overall degree of inter-LA migration is relatively low, with around 4% of families moving LA each year after children turn 5. Since out treatment is defined by the child's LA of residence at the time of admission, this provides further reassurance that inter-LA migration is not a major source of measurement error in our context.

5 Sure Start's effects on overall hospitalizations

5.1 Main estimates

Table 1 reports the estimates of the effect of a one-center (per thousand children) increase in access to Sure Start on hospitalizations for any cause between the ages of 1 and 15. These effects are estimated separately from 15 regressions (one for each age of admission). Figure 3 plots these estimates re-scaled by the baseline probability (measured in the cohort born in 1996) of any hospitalization at the corresponding age to enable comparison of relative effects across ages.

These results show that, during the earliest years of life, an increase in Sure Start coverage resulted in an increase in hospital admissions. In particular, an additional center per thousand children raises the probability of any hospitalization at age 1 in a cell by 2.6 percentage points, a 10% rise relative to the pre-Sure Start baseline (when 26% of LSOA-sex-quarter of birth cells had at least one hospitalization). This translates into approximately 6,700 additional yearly hospitalizations.[23]

However, as Figure 3 shows, these early increases in hospitalizations are followed by substantial decreases in the probability of admission through childhood and early adolescence. Once

[23] To compute the number of yearly additional or averted hospitalizations engendered by the presence of an additional center per thousand children, we multiply the estimates of parameter δ^{ya}, as defined in model 1 and presented in Table 1, by the number of observations per year, which is 262,528 (32,816 LSOAs by 4 quarters and 2 sexes).

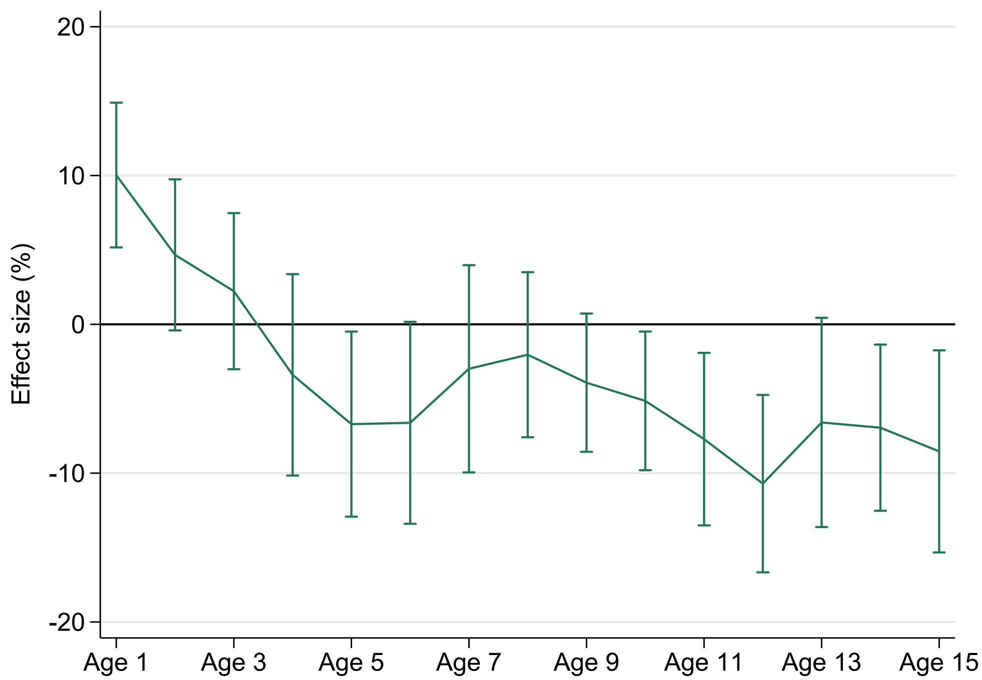
Table 1: Effect of an increase in Sure Start coverage on probability of hospitalization for any cause

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
SS Cov	0.0256***	0.0095	0.0040	-0.0057	-0.0109*	-0.0095	-0.0038	-0.0024	-0.0044	-0.0055*	-0.0084**	-0.0125***	-0.0081	-0.0091**	-0.0120**
	(0.0075)+++	(0.0063)	(0.0057)	(0.0069)	(0.0061)	(0.0059)	(0.0053)	(0.0039)	(0.0032)	(0.0031)	(0.0038)+	(0.0042)+++	(0.0052)	(0.0045)	(0.0058)
Baseline mean	0.2552	0.2044	0.1791	0.1687	0.1623	0.1438	0.1260	0.1160	0.1125	0.1078	0.1089	0.1172	0.1229	0.1311	0.1410
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: The table shows coefficients from regression analysis at each outcome age. Observations are cells defined by the LSOA, quarter-year of birth, and sex. The model regresses an indicator for any hospitalization in a cell on Sure Start coverage, the population at the relevant age in the LSOA, an indicator for female and fixed effects for quarter–year of birth and for the LSOA of residence. Sure Start coverage SS_{qd} is defined as the number of centers per thousand children aged 0–4 in the local authority for outcomes at age 5 and older, and as the number of centers per thousand children aged 0-4 that were open in the LA d when the child was aged $a-1$ for outcomes at ages $a = 1, ..., 4$. The baseline mean (3rd row) is measured in 1996. 'Earliest' and 'latest' cohorts refer to the first and last birth cohort included in each regression. Standard errors are shown in parentheses clustered by LAD. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; $+$, $++$ and $+++$ indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure described in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

children turn 5 and stop being age-eligible to use Sure Start services, the overall impact on hospitalizations becomes consistently negative, with larger impacts during the first few years of schooling (ages 5-6) and then from age 10 onward. Exposure to an additional center per thousand children at ages 0-4 averts around 7% of hospital admissions at age 5, 8% by the end of primary school at age 11, and 8.5% by age 15 (the final age we study). This represents around 2,860 fewer yearly hospitalizations at age 5 and over 13,150 prevented hospitalizations of 11- to 15-year-olds each year. Table 1 also indicates whether the estimates are still significant after adjusting inference to multiple hypothesis testing: the increase in admissions among infants and the reductions at ages 11 and 12 survive this adjustment.

Figure 3: Effect of an increase in Sure Start coverage on probability of any hospitalization in the neighborhood, rescaled by baseline probability



Note: Effect sizes are constructed by rescaling the estimates by the pre-Sure Start baseline probability of a hospitalization at each age. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

5.2 Sensitivity Analyses

As discussed in subsection 4.3, our difference-in-differences design relies on the assumption that cohorts' exposure to Sure Start is uncorrelated with time-varying unobservable shocks to hospitalizations. In this section we further challenge the validity of our identifying assumption by presenting a series of alternative specifications and placebo analyses.

5.2.1 Validity of identification strategy

Differential trends We first augment the model to allow for differential trends in hospitalizations across Local Authorities in two different ways. Specifically, we first estimate LA-specific linear time trends in the pre-treatment period, i.e. using cells on cohorts for which $SS_{dq} = 0$. For each LA, we obtain a slope estimate λ_d. We then linearly extrapolate this LA-specific pre-treatment time trend for all the cohorts in the sample and include this estimated trend as a control in our main model (equation 1). These estimates are presented in Appendix Figure A.6 and are similar to our main estimates of Figure 3.[24]

The second way we probe the robustness of our results to differential LA trends is by including in the benchmark model interactions between a cohort trend and the baseline characteristics that we showed affected the rollout of Sure Start (1998 deprivation levels, teen conception rate and incidence of low birth weight). Appendix Table A.6 shows that our main estimates remain largely unchanged in these specifications, whether we use a linear or quadratic cohort trend.

Confounding variables We also estimate a version of model (1) that includes a wide range of time-varying local area characteristics, including measures of other public services that changed over this period. We conduct two versions of this robustness check, measuring these local area

[24] We estimate LA-specific trends using data from before the implementation of SS, and we obtain a slope estimate λ_d for each LA. We then extrapolate the pre-expansion time trends to the post-reform period as follows (see also (Bhuller et al., 2013)):

$$D^{ya}_{sql(d)} = \delta^{ya} SS_{dq} + \pi^{ya}_{l(d)} + \gamma^{ya}_q + \delta\widehat{\lambda_d}t + \epsilon^{ya}_{sql(d)}, a = 1, ..., 15.$$

characteristics either in the child's year of birth or contemporaneously in the year that the outcomes are measured. The former version tests for confounding variables that were tied to the Sure Start rollout and that may have influenced children's early health, such as the teen conception rate or the health service offer. The latter specification tests whether our estimates are confounded by a correlation between the rollout of Sure Start and subsequent changes in local characteristics or the local service offer, for example from policymakers seeking to 'follow up' early intervention with later services.

Figure 4 shows that both of these robustness checks yield very similar results to our main estimates (though there are differences at age 15). We interpret the robustness of our results to this wide range of local characteristics as evidence that the internal validity of our research design is not compromised by most plausible confounders.[25]

Placebo checks We subject our results to a placebo check, by considering any admissions due to congenital chromosomal defects before age 1: since these are genetic conditions, they cannot be plausibly affected by Sure Start. We therefore expect our estimates to show no impact of Sure Start on admissions for these conditions. Indeed, Table A.7 shows that increasing SS coverage is unrelated to changes in the likelihood of hospitalization in early life due to congenital chromosomal defects, with precisely estimated null effects.

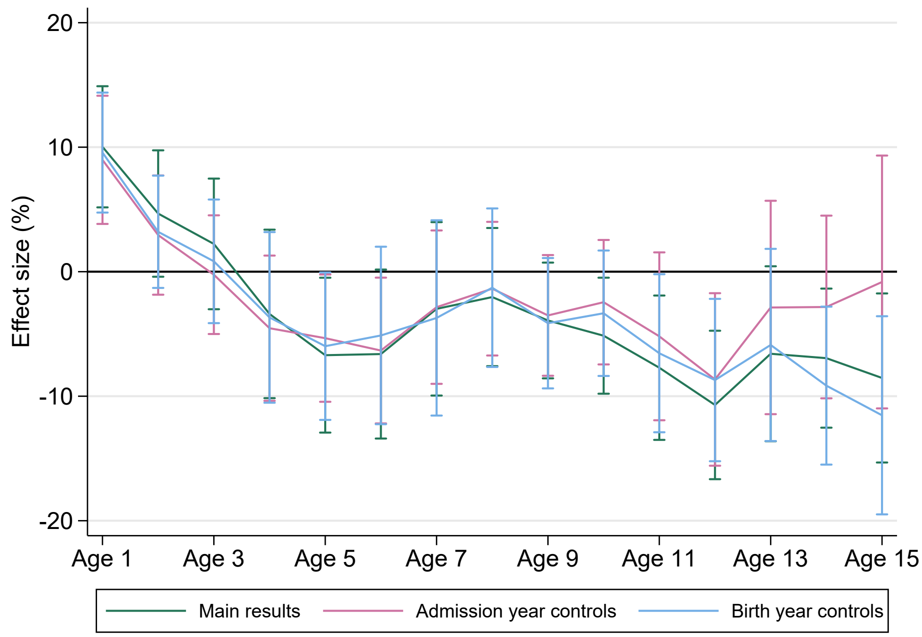
5.2.2 Sensitivity of the results to sample selection

As indicated at the bottom of Table 1, the sample of analysis is not constant across outcomes. This is because we have sought to maximize our sample at each age, within constraints imposed by the data and the need for comparability across cohorts.[26] In order to check that our results

[25]This specification include a variable measuring the availability of free funded childcare places for 3- and 4-year-olds, which were rolled out across areas and years as part of the 'free entitlement' policy. Given the results of this specification, there is no reason to expect that the policy is confounding the effects of Sure Start. Nevertheless, we re-estimate the main model where we now only include one additional variable: the proportion of 3- and 4-year-olds with a funded place measured at yearly level and at the Local Authority. The impacts of Sure Start coverage are mostly unchanged. The coefficients on the free entitlement variable are statistically insignificant for the most part, suggesting that the expansion of this childcare subsidy did not affect hospitalizations (results available upon request).

[26]Many younger cohorts are not yet old enough to have data for hospitalizations at later ages. Further, since the inpatient data are only collected from 1997, some older cohorts will not be observed at younger ages.

Figure 4: Effect of an increase in Sure Start coverage on probability of any hospitalization, rescaled by baseline probability: Robustness to inclusion of time-varying controls



Note: Figure shows coefficients from separate regressions for each outcome age. Coefficients are rescaled by the baseline (1996) mean for each age. Specification including time-varying controls contains controls for: the teenage conception rate; the share of births with low birth weight; the total period fertility rate; the LA population density; the share of primary school students with English as an Additional Language; the rate of Children Looked After among infants and among children aged 1-4; the Jobseeker's Allowance receipt rate; the number of GPs per capita in the LA; the number of JobcentrePlus per capita in the LA; and the take-up rate for funded childcare places for 3- and 4-year-olds in the LA. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start. Area characteristic sources are in Appendix Table B.1.

are not driven by changes in the composition of the sample across ages, we re-estimate our main specification on two common cohorts. The first common cohort covers ages 1 to 4 and uses data for children born between April 1996 and December 2006. As Figure A.7 shows, the results on this early years cohort are virtually identical to our main estimates. The second common cohort covers ages 11 to 15 and uses data for children born between January 1993 and March 2002. Figure A.8 shows that the results on the common teen cohort are statistically indistinguishable from our main estimates.

5.2.3 Non-linearity in treatment effects

Our main model assumes a linear effect of Sure Start coverage on children's hospitalizations. It may however be possible that the effect is non-linear, for example if there needs to be a critical mass of children exposed to Sure Start to start picking up effects on hospitalizations or if effects arise only once families have access to several centers in their vicinity. We explore this possibility by re-estimating our model to distinguish between no Sure Start coverage, medium coverage (fewer than 0.25 centers per thousand children), and high coverage (more than 0.25 centers per thousand).[27] Figure A.9 shows that the impacts of high coverage are of greater magnitude than the impacts of medium coverage, but other than at age 1, the impacts of medium and high coverage are statistically indistinguishable. This is reassuring about the validity of our linear specification.

5.3 Intensive margin

Our main specification captures the effect of an increase in Sure Start coverage on the probability of any hospitalizations occurring in the neighborhood-cohort-sex cell. In Panel A of Table A.8, we also report the estimates of the effect of Sure Start coverage on the number of hospital admissions in the cell. Like in our main specification, point estimates are positive in early ages and then turn negative around age 4. However, only impacts at early ages are statistically significant, which is unsurprising given that, especially at later ages, there is not a lot of variation in the number of admissions across cells.

5.4 Robustness to heterogeneous treatment effects

Recent work has emphasized that the estimand recovered in the linear model with Two Way Fixed Effects (TWFE), as we employ here, is a weighted sum of the average treatment effects (ATE) in each group and period, with weights summing to one but some possibly negative (Borusyak, Jaravel and Spiess, 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021).

[27] The cut-off point between medium and high coverage is approximately the median coverage among those with positive coverage.

Negative weights arise when groups that are treated earlier are used as controls for groups that are treated later, and hence are more likely to be assigned to long-run ATEs. As discussed in these papers, if treatment effects are heterogeneous across group and period, the presence of negative weights creates an issue in that the treatment parameter recovered in the standard TWFE model can be negative even if all ATEs are positive. Moreover, the TWFE estimand might not necessarily be the aggregation scheme that researchers might find to be most relevant to focus on (Callaway and Sant'Anna, 2020).

In the context of Sure Start, heterogeneity of treatment effects across groups and time is a possibility. Moreover, because we exploit the staggered rollout of Sure Start, it is possible that our estimand puts negative weights on some treatment effects. Although we rely on a continuous treatment measure, and not a discrete treatment as most papers in this literature consider, Callaway, Goodman-Bacon and Sant'Anna (2021) do show that the same type of issues would apply in our case. Indeed, in all our models, the proportion of negative weights is around 50%, though their sums varies between -0.17 and -0.53 (see Table A.9).

As no alternative estimator for continuous treatments has yet been developed, we assess the severity of these issues by discretizing our treatment variable and comparing the binary treatment effect estimates we obtain using the TWFE estimator with those we obtained using the efficient imputation estimator proposed by Borusyak, Jaravel and Spiess (2021) (BJS henceforth).[28] We consider three different binary treatment variables defined as indicators for whether SS_{dq} is above 0, 0.1 and 0.25. The results comparing the effects of these binary treatment effects are reported in Appendix Table A.10. In general, the TWFE and the BJS estimates are very similar to each other, suggesting that negative weights are unlikely to be an important issue in our context. For infants, Table A.10 shows the BJS estimator yields slightly larger impacts on hospitalizations than the TWFE estimator, especially when the treatment variable is defined as there being any coverage

[28] We choose the imputation estimator of Borusyak, Jaravel and Spiess (2021) over the ones proposed by de Chaisemartin and D'Haultfœuille (2020) and Callaway and Sant'Anna (2020) because it is more efficient than the others under heteroskedasticity. The gain in efficiency comes from the fact that the imputation estimator of BJS uses all non-treated periods to impute the counterfactual outcome for each group, while the alternative estimators only use the one period before the group becomes treated as counterfactual.

($SS_{dq} > 0$). From age 10 onward, the TWFE and the BJS estimators yield similar estimates, especially for treatment indicators of 0.1 or 0.25 centers per 1,000 under 5 children, though we note that the BJS model often yields more precise estimates than the TWFE (at least in the case of binary treatment effects).

6 Mechanisms

Having shown in section 5 that greater access to Sure Start increased the hospitalizations of infants but reduced admissions among older children, we now turn to discussing the mechanisms underlying these impacts.

6.1 Program objectives, service provision and cause–specific hospitalizations

We start by shedding light on the aspects of this multi-faceted program that were most effective in driving the effects on overall hospitalizations reported so far. Because there is no data linking measures of health or hospitalizations with information on Sure Start services offered to or taken up by families, we proceed in two ways. First, we analyze the impact of Sure Start on the profile of hospitalizations for different causes. As we discussed in section 3, for most Sure Start services we can hypothesize how an effective service might impact the profile of hospitalizations for different causes. We can use results of this analysis to infer what aspects of the program may have been particularly effective. However, as set out above, the impact on hospitalizations of services targeting parental employment is ambiguous. Thus, we turn to use survey data from the UK Labour Force Survey (LFS) to present direct evidence of the impact of greater access to Sure Start on parental employment.

Cause-specific hospitalizations Following the discussion in section 3, we estimate the impact of greater access to Sure Start on hospitalizations for conditions that are most likely to have been affected by Sure Start: preventable conditions, infectious illnesses, external causes (accidents, injuries and poisonings), and (among adolescents) mental health.[29] We measure preventable conditions as Ambulatory Care Sensitive (ACS) conditions, which include chronic conditions that can typically be managed outside of hospital (e.g. asthma); acute conditions where serious illness could have been prevented by early intervention (e.g. gangrene); and conditions that arise from vaccinable diseases (e.g. measles).[30] We then re-estimate our main model (equation 1) in the same sample used for the benchmark results reported in section 5. As before, we present results graphically in Figure 5 and report estimates and p-values adjusted for multiple hypothesis testing in Appendix Table A.11 .

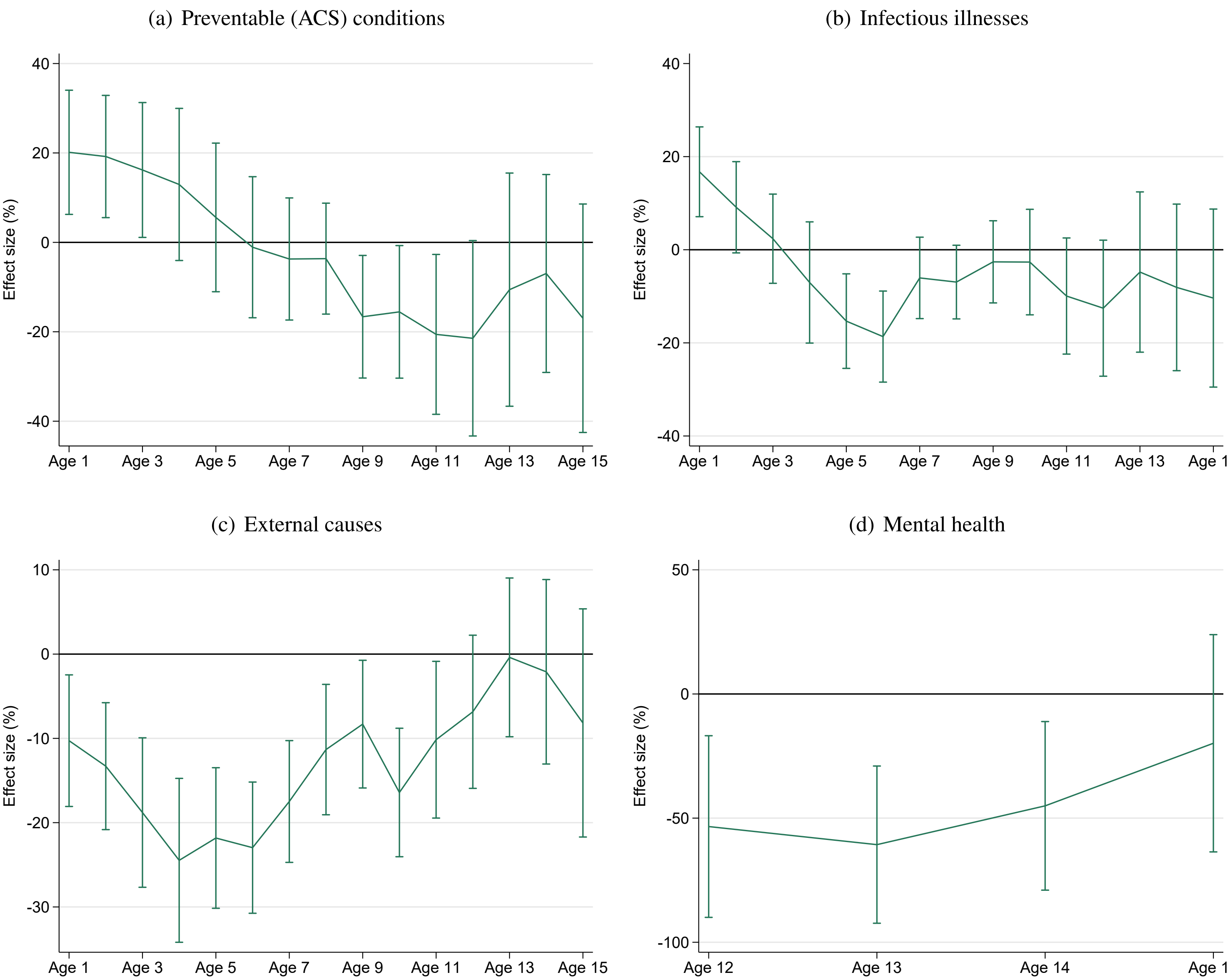
Starting with preventable (ACS) conditions, Figure 5(a) shows that access to Sure Start substantially increases hospitalizations for these conditions at younger ages, with a 20% increase over baseline levels at age 1. However, as children age, greater access to Sure Start instead reduces ACS admissions, with a 20% reduction over baseline levels by age 11. This pattern is consistent with Sure Start providing information and signposting to help parents identify and learn how to manage their child's conditions earlier in life, thereby reducing hospitalizations later on.

Figure 5(b) shows Sure Start's impacts on hospitalizations for any infectious illness (which include infectious and parasitic diseases and respiratory illness). We find that greater access to Sure Start substantially increases hospitalizations for infectious illnesses in infancy (see also van den Berg and Siflinger (2022)); however, there are significant and substantial falls in hospitalizations (of up to 18% of the baseline) at ages 5 and 6, just after children age out of Sure Start eligibility and start school. In line with the discussion in section 3, the results presented here are consistent with exposure to pathogens through Sure Start activities such as childcare: children who are more exposed early in life are initially more vulnerable to infectious illness, but then build up a stronger immune response which protects them compared to their peers when the entire cohort enters school. These effects then fade out in the longer term, as the start of universal schooling sees other children's immune systems 'catch up'.

[29]Hospital admissions in the HES data can have up to 20 causes, recorded via ICD-10 codes. We classify admissions based on the primary diagnosis recorded; however, our results are similar when we instead look for any diagnosis matching the criteria.

[30]See Blunt (2013) for a full list of ICD-10 codes that are included in this definition.

Figure 5: Effect of an increase in Sure Start coverage on probability of hospitalization for specific causes, re-scaled by baseline probability



Note: Figure shows coefficients from separate regressions for each outcome age. Coefficients are re-scaled by the baseline (1996) mean for each age. Vertical bars indicate 90% confidence intervals. Cause-specific results are based on the primary diagnosis at the time of admission. See Blunt (2013) for a list of all relevant ICD-10 codes included in ACS conditions. Infectious illnesses are composed of infectious and parasitic diseases (ICD-10 groups A and B) and respiratory illnesses (ICD-10 group J). External admissions include ICD-10 codes in groups S, T, V and Y. Mental health admissions relate to ICD-10 codes beginning with F.

Next, we turn to hospitalizations for external causes.[31] Figure 5(c) shows that there is a very

[31] Those correspond to ICD-10 groups S, T, V and Y

large, significant decline in hospitalizations for external causes at almost all ages we consider. These estimates are always negative, and even at the youngest ages the probability of an externally caused hospitalization falls by 10% or more with greater access to Sure Start. At younger ages, these results offset some of the increase in hospitalizations due to infectious illnesses.

To understand the mechanisms underlying those impacts, we further analyze the impacts of Sure Start on different categories of external causes. Table A.12 shows that greater access to Sure Start significantly reduces poisonings from ages 1 to 3, consistent with information about or direct provision of safer environments for young children. However, by far the main driver of reductions in hospitalizations for external causes is a reduction in injuries, which decline with greater access to Sure Start during almost all years in childhood.[32]

The magnitude and persistence of the effect on injuries is also consistent with sustained impacts on children's emotional and behavior development. Indeed, several studies report a correlation between children's behavioral issues (e.g. hyperactivity and aggressive behavior) and hospitalizations for injuries (e.g. Hoare and Beattie, 2003). In Appendix D, we complement this evidence using data from a nationally representative cohort of children born in 2000-2002 (the Millennium Cohort Study) and show that having fewer externalizing behavior problems (such as aggression or hyperactivity) is correlated with a reduced probability of injury in middle childhood and early adolescence, even conditional on a very rich set of demographics and family circumstances. The effect of Sure Start on reducing injuries could also reflect the effect of the program on reducing child maltreatment (through parenting and broader family support provided by the program).[33]

Lastly, we look at the impact of Sure Start on children's mental health-related admissions. Recorded mental health hospitalizations among young people are extremely rare (for example, among children they occur in just 0.02% of cells). As a result, we focus our analysis on outcomes among teenagers (ages 12 to 15). In Figure 5(d) (and Appendix Table A.13), we show that addi-

[32]Injuries (ICD-10 groups V and Y) account for between 70 and 80% of external admissions; most of the rest are accounted for by poisonings (codes T15-T98).

[33]Reductions in hospitalizations for injuries are commonly interpreted in the home visiting literature as signs of reductions in child maltreatment (Kitzman et al., 1997). Previous research has identified a subset of conditions that can be used as proxies for potential maltreatment, but the incidence of these is too low to reliably estimate Sure Start's impacts on these outcomes (González-Izquierdo et al., 2010) directly (results available upon request).

tional access to Sure Start significantly decreases in mental health-related admissions at ages 12 to 14. We interpret this result as another piece of evidence pointing to Sure Start improving the socio-emotional development of children.

Of course, there are limitations in using hospitalization data to look at mental health: hospital admissions for mental health conditions are an extreme outcome and do not capture young people who are receiving services in the community, through their schools or through non-hospital providers. Moreover, previous work has also raised concerns about the accuracy of mental health diagnosis coding, especially for conditions such as depression or anxiety (Davis, Sudlow and Hotopf, 2016). For these reasons, in subsection 6.2, we complement the results presented here with analysis of survey data looking at the impact of Sure Start on self-reported measures of socio-emotional development among adolescents.

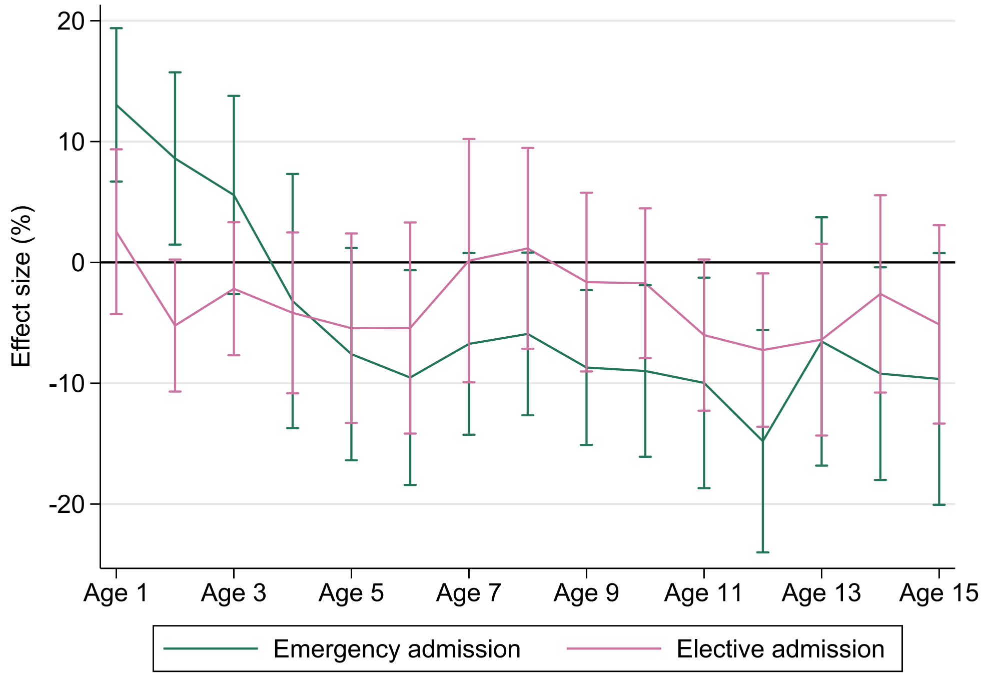
Parental employment In addition to their focus on children's health and development, Sure Start centers also brought together existing services to support parental (especially maternal) employment. In section 3, we discussed the several channels through which an increase in parental employment resulting from Sure Start could affect children's hospitalizations. In order to tease out whether this is likely to be an important mechanism underlying our results, we use another dataset, the UK Labour Force Survey, to directly estimate whether increased Sure Start access had an effect on maternal labor market outcomes. These data have a quarterly frequency and a rotating panel structure at the household level (similar to the Current Population Survey in U.S.). As a result, we need to adapt our estimation strategy to account for the smaller sample size and to take advantage of the longitudinal dimension of the data. Appendix F describes the data and estimation framework and presents the results. We find no robust evidence that Sure Start affected maternal labor supply, either when children were aged 0 to 4 or later on. We conclude from this analysis that it is unlikely that the effects we observe on children's hospitalizations are driven by an increase in maternal employment (and family income).

6.2 Health vs. healthcare utilization

Children's hospitalizations are a measure of healthcare utilization, which reflects both children's underlying health and families' propensity to take up healthcare (holding health constant). Given the nature of services offered by Sure Start, the program may have affected both margins. While the results on cause-specific admissions presented above suggest that Sure Start may have increased families' propensity to take their children to hospital during infancy, the reductions in hospitalizations in childhood and adolescence strongly suggest that the program improved children's underlying physical and mental health. In this section, we provide further evidence, based on administrative and survey data, that the program did more than change the take-up of families' hospital care and that, indeed, it had a positive impact on children's health in the medium term.

Evidence based on administrative data Using our benchmark specification in the hospital admissions data, we examine the impacts of greater access to Sure Start on two additional outcomes related to hospitalizations. First, we examine whether effects are heterogeneous across the two possible routes through which patients can be admitted to a publicly funded hospital: via the emergency room or via the elective route (which in England can only be accessed following a referral by a NHS General Practitioner). The results are shown in Figure 6 (underlying coefficients are presented in Appendix Table A.14). The change in admissions resulting from an increase in Sure Start coverage is clearly driven by Sure Start's impacts on emergency admissions, with null effects on elective admissions for most ages. This is consistent with Sure Start affecting the incidence of illness or injury as shown above, not just families' propensity to seek health care for underlying or longer-term conditions.

Figure 6: Effect of an increase in Sure Start coverage on probability of any hospitalization, re-scaled by baseline probability: Emergency and elective admission routes



Note: The figure shows coefficients from separate regressions for each outcome age. Coefficients are re-scaled by the baseline (1996) mean for each age. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

Using the same data, we also study the impact of greater access to Sure Start on the average number of days children and adolescents spent hospitalized. These estimates are presented in Panel B of Table A.8. The estimates show that greater access to Sure Start did not have any effect on the average length of hospital stay of infants, indicating that the program did not make infants more severely ill than they would have been otherwise. The increase in the probability of being hospitalized at this age (which we have shown is driven by increases in hospitalizations for ACS and infectious diseases in Figure 5) is therefore likely to reflect an increase in the number of sick children (as a result of greater contact with other children) and/or an increase in families' propensity to take their children to hospital (as a result of the program's provision of information on appropriate healthcare for infants).

In contrast, at age 10-11, the reduction in the proportion of hospital admissions reported in section 5 is accompanied by a reduction in their average length of stay. These results suggest that the reduction in hospitalizations seen from age 11 reflect an improvement in children's health. This is further supported by the fact that the reduction in admission during early adolescence is due to the reduction in hospitalizations for external conditions, in particular injuries and fractures, which need timely treatment.[34]

Evidence based on survey data We complement this analysis by estimating the impact of greater access to Sure Start on several direct measures of health using the UK Household Longitudinal Study (UKHLS). The UKHLS covers a representative sample of around 40,000 households, with annual interviews starting in 2010. Adolescents in the household (aged 11 to 15) self-complete a dedicated survey covering topics including self-reported health and mental health (as assessed by the Strengths and Difficulties Questionnaire). Because of the relatively small sample size of adolescents born in the window of interest to study the effects of Sure Start on these measures, we cannot control for small neighborhood (LSOA) fixed effects and hence cannot implement the same specification as the one we implement with administrative hospital data. Instead, we exploit the longitudinal nature of the UKHLS to estimate a family fixed effects specification exploiting variation in exposure to Sure Start between siblings (see subsection E.1 for details about the exact specification that we estimate and the results).

As reported in subsection E.1, young people who were more exposed to Sure Start than their siblings were much more likely to report being in 'very good' or 'excellent' health in secondary school. It also improved young people's mental health: adolescents who were more exposed to Sure Start in their childhood displayed significantly fewer socio-emotional difficulties than their siblings, and especially fewer internalizing behaviors. This provides reassuring evidence that the marked declines in hospitalizations among adolescents is related to improvements in underlying physical and mental health.

[34]Relying on LSOA-level mortality records, we do not find effects of Sure Start on infant or child mortality rates (see Appendix E for details).

7 Impact heterogeneity by gender and deprivation

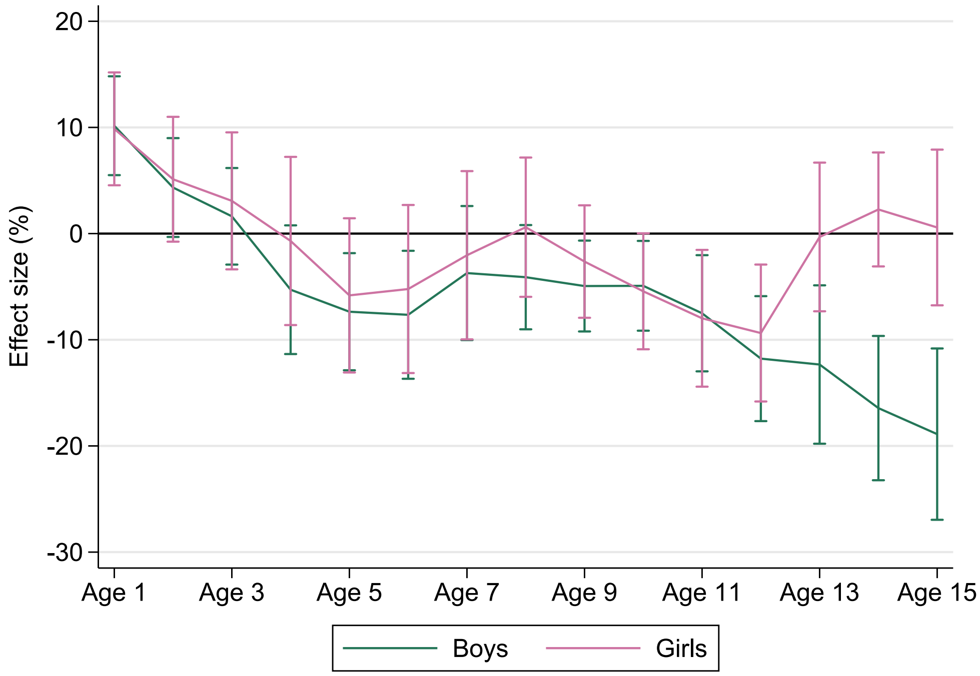
The literature evaluating early childhood interventions often report the presence of heterogeneous impacts across different groups of children. We explore whether impacts of Sure Start on hospitalizations are heterogeneous by gender and by areas with different levels of deprivation. The latter dimension is particularly relevant given that, to date, the evidence available on the health effects of early childhood interventions offering both preschool education and family support services is based on interventions targeted at disadvantaged children.

7.1 Heterogeneity by gender

Figure 7 (and the point estimates in Table A.15) shows how the effects of Sure Start on all-cause hospital admissions vary between girls and boys. While the profile of effects is fairly similar for girls and boys up to age 10, during adolescence the impacts diverge: there is no impact on girls in their teen years, but the impact on boys grows steadily. By age 15, an additional Sure Start center per thousand children during the first five years of life reduces the probability of hospitalization among boys by 20%, with no effect among girls.

The greater impacts on boys in adolescence are consistent with the results of other early childhood intervention evaluations, such as the Abecedarian program (Conti, Heckman and Pinto, 2016), Head Start (Carneiro and Ginja, 2014) and the Boston Preschool program (Gray-Lobe, Pathak and Walters, 2021). Interestingly, in the case of Sure Start, we find similar impacts for boys and girls early on, suggesting that the gender difference in impacts during adolescence is unlikely to be due to differences in the take-up of services. An analysis of gender-specific effects on hospitalizations for different causes reveals that the gender difference in the program impacts is entirely driven by the greater impact of Sure Start reducing hospitalizations for injuries for boys (see Figure 8). Behavioral problems being more frequent among boys than girls (Bertrand and Pan, 2013), this finding could be further indication that an important channel through which the program worked was by improving children's behavioral development. This argument is bolstered by the pattern of take-up of Sure Start services: as we show in Table A.3, the parents of boys are more likely to take up parenting support and parent-child services, but there are no gender gaps in the use of other Sure Start services.

Figure 7: Effect of an increase in Sure Start coverage on probability of any hospitalization, rescaled by baseline probability: Differences by gender

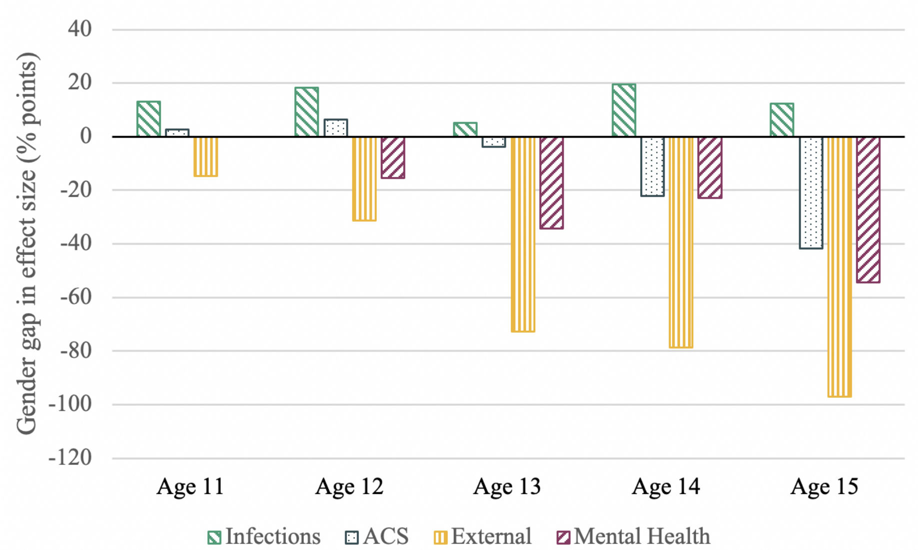


Note: The Figure shows coefficients from separate regressions for each outcome age, with Sure Start treatment interacted with gender. Coefficients are rescaled by the gender-specific baseline (1996) mean for each age. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

7.2 Heterogeneity by level of deprivation

Sure Start started as an intervention targeting highly disadvantaged areas, but the program was universalized from 2004 onward. Many large-scale early childhood interventions have been found to disproportionately benefit more disadvantaged populations (see Almond, Currie and Duque (2018) for a review), and we now turn to study whether the program's impacts varied by socioeconomic status.

Figure 8: Gender gap in the Effect of an increase in Sure Start coverage on probability of cause-specific hospitalizations



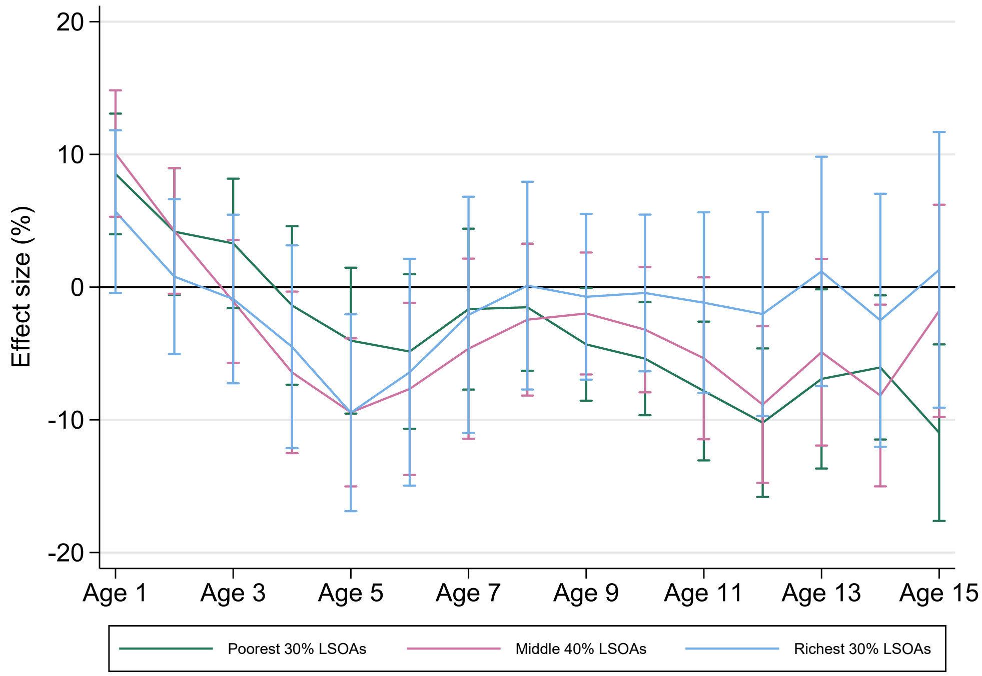
Note: The Figure shows the percentage point difference the estimated effect size of Sure Start on the probability of hospitalization between boys and girls. The difference in effect size between both genders is statistically significant at the 90% level at ages 11-12 for infections, age 15 for ACS and ages 11-15 for external. Table A.16 and Table A.17 display the original cause-specific point estimates and p-values by gender for ages 1-15.

Because we do not have information on family income or parental education in the hospitalization data, we rely on the level of deprivation of the neighborhood of residence.[35] In particular, we allow for heterogeneity of Sure Start effects for the three groups of neighborhoods (defined at the LSOA level): the 30% most deprived neighborhoods, the 30% least deprived neighborhoods, and those in the middle of the distribution of disadvantage. As Figure 9 and Table A.18 illustrate, the increase in admissions among infants detected for the whole sample in Figure 3 is driven by those residing in neighborhoods falling into the poorest 30% of the deprivation distribution. From ages 10 to 15, the drop in hospital admissions attributed to Sure Start is even more concentrated in the areas with the highest levels of disadvantage, with imprecise impacts at the middle of the

[35]To classify neighborhoods into these three groups, we use the 2004 Index of Multiple Deprivation (IMD). The IMD 2004 contains seven domains of deprivation: Income deprivation; employment deprivation; health deprivation and disability; education, skills and training deprivation; barriers to housing and services; living environment deprivation; and crime. Each domain contains a number of indicators, which are aggregated by principal component analysis. The IMD is the government's measure of small area deprivation and its 2004 version was the measure used to classify areas into the different stages of the rollout of Sure Start Children's Centres.

distribution (and zero effects at the top of the distribution).[36]

Figure 9: Effect of an increase in Sure Start coverage on any hospitalization, rescaled by baseline probability: Differences by area deprivation



Note: Figure shows coefficients from separate regressions for each outcome age, with Sure Start treatment interacted with the three disadvantage categories. Coefficients are rescaled by the deprivation-specific baseline (1996) mean for each age. Vertical bars indicate 90% confidence intervals.

The stronger impacts of Sure Start among more deprived neighborhoods could be a result of a number of different and non-mutually exclusive factors having to do with service quality, the quality of the counterfactual environment, and service take-up. On the one hand, there is ample evidence that disadvantaged children grow up in less safe and stimulating environments and that disadvantaged parents make less use of healthcare (Currie, 2006). Disadvantaged families may therefore have had more scope to benefit from the information and services to support parents that

[36]As discussed in section 2, the rollout of Sure Start started in the poorest areas and progressively expanded into richer areas. Given that we use hospitalization data until 2017, the sample we use to measure impacts on hospitalizations at age 15 only includes cohorts born up until 2003 (cohorts born until 2004 for age 14, etc.). This means that there may be less variation in exposure among the least deprived cohorts than among the most deprived cohorts to identify the effects of Sure Start on hospitalizations during adolescence. We check whether this is the case by plotting the variation in each of the three subgroups in Appendix Figure A.10, which shows that there is still a lot of variation in exposures to Sure Start for relevant cohorts in the middle and richest neighborhoods.

Sure Start provided. The services offered to families in more deprived areas may also have been of higher quality or intensity than those offered in less deprived areas. As mentioned in section 2, Phase 1 and Phase 2 SSCCs (in the 30% most deprived areas) had more stringent requirements in terms of the qualifications of the childcare staff than Phase 3 SSCCs. If staff qualifications are indeed an important input for the production of high-quality childcare (Sylva et al., 2010), the rollout of Sure Start may have meant that children in more deprived areas accessed childcare of greater quality relative to the counterfactual than children in less deprived areas. Finally, the stronger impacts of Sure Start in deprived areas may also reflect a higher take-up of services by more disadvantaged families. Survey data on service offer and service take-up presented earlier in subsection 3.3 does suggest that this could be the case for parenting support services, which evidence presented so far suggests was a particularly important component of the program (see Figure A.11 and Sammons, Goff and Smith (2015)).

8 Conclusion

Early childhood interventions that offer disadvantaged children preschool education and family support services in the US show consistent and long-lasting impacts on the health of their participants. However, much remains to be known about whether these impacts translate into less targeted programs, especially in contexts with more generous safety nets. Answering this question is highly relevant to inform current policy efforts to integrate early years services in a number of OECD countries and expand investments in early childhood interventions in the U.S. and continental Europe as a way to decrease health expenditures and reducing health inequalities.

The contribution of this paper is to show that a universal integrated early childhood intervention can deliver significant and long-lasting health benefits, even in a context with free healthcare. We exploit a unique social experiment - the rollout of Sure Start, an area-based program offering health services, early learning and childcare, parenting support and parental job search assistance to families with a child under 5 in England. We use administrative data on the universe of hospital admissions in publicly funded hospitals and unique data about when and where every Sure Start centers opened between 1999 and 2010. Despite the national importance of Sure Start in England, evidence about its impact is scarce. This paper presents the first robust quasi-experimental evaluation of the benefits of accessing Sure Start centers for children.

We find that greater access to Sure Start increased hospitalizations during infancy, but subsequently reduced them during childhood and adolescence. Among infants, having access to an extra center per thousand children increased the probability of a hospitalization in the neighborhood cohort by 10% of the baseline at age 1. Once children turn 5 and stop being age-eligible to use Sure Start services, the overall impact on hospitalizations becomes consistently negative. Exposure to an additional center per thousand children under five reduces hospital admissions by 8% at the end of primary school (age 11) and by 8.5% at age 15.

The profile of impacts on overall hospitalizations masks substantial heterogeneity in the profiles of impacts on hospitalizations for specific causes. At the youngest ages, greater access to Sure Start increased hospitalizations, driven mainly by an increase in infectious illnesses. The increase was partly offset by a fall in hospitalizations from external causes and poisonings in the early years. Later, during early primary school, hospitalizations related to infectious illness fell. In later primary school years and early adolescence, we also observe statistically fewer admissions to hospital for mental health reasons.

These patterns are consistent with Sure Start improving children's health and other dimensions of development through a number of key mechanisms: providing parents with greater information about children's health and healthcare; strengthening children's immune systems; and improving children's behavioral and emotional development, by improving parenting practices and/or providing high-quality childcare. Coupled with survey-based evidence on the program's impact on self-reported health and mental health measures, these results suggest that early childhood interventions focusing on these channels can deliver lasting health benefits, even in contexts with universal free health care.

Overall, these results reflect the importance of integrating health services with early education and childcare and parenting services to promote child development in a holistic way. The persistence of impacts into middle childhood and adolescence is driven by the reduction in hospital admissions for injuries and mental health. These impacts are likely to reflect an improvement in children's behavioral and emotional development, which we confirm with our analysis of survey data. Together, these results strongly speak to the importance of cross-productivities between different domains of development (Cunha, Heckman and Schennach, 2010; Cunha et al., 2006).

A simple cost-benefit analysis shows that the financial benefits from reduced hospitalizations offset approximately 31% of the provision cost of Sure Start (see Appendix G). This figure should be interpreted as a lower bound of the program benefits because it disregards impacts on other outcomes that could have also been affected by Sure Start. Nevertheless, our results on health suggest that the overall effectiveness of the intervention might have come despite, rather than because of, its universality. Indeed, impacts are concentrated in the 30% most disadvantaged neighborhoods. In line with evaluations of other universal preschool programs, some form of targeting might therefore have been desirable to reach a higher value for money. In contrast with HeadStart, Sure Start was area-based, which may be an attractive alternative to individual means-testing to potentially reduce individual stigma associated with attending a targeted program.

References

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4).

Anders, John, Andrew Barr, and Alexander A. Smith. 2022. "The Effect of Early Childhood Education on Adult Criminality: Evidence from the 1960s through 1990s." *American Economic Journal: Applied Economics*, forthcoming.

Avellar, Sarah A, and Lauren H Supplee. 2013. "Effectiveness of home visiting in improving child health and reducing child maltreatment." *Pediatrics*, 132(Supplement 2): S90–S99.

Bailey, Martha, Shuqiao Sun, and Brenden Timpe. 2020. "Prep School for Poor Kids: The Long-Run Impact of Head Start on Human Capital and Productivity." Working Paper.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. "Universal Child Care, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy*, 116(4): 709–745.

Barr, Andrew, and Chloe Gibbs. 2022. "Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood." *Journal of Political Economy*, forthcoming.

Bertrand, Marianne, and Jessica Pan. 2013. "The Trouble with Boys: Social influences and the gender gap in disruptive behavior." *American Economic Journal: Applied Economics*, 5(1): 32–64.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249–275.

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): 1101–1157.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad. 2013. "Broadband Internet: An Information Superhighway to Sex Crime?" *The Review of Economic Studies*, 80(4): 1237–1266.

Bitler, Marianne, Hillary Hoynes, and Thurston Domina. 2014. "Experimental Evidence on Distributional Effects of Head Start." NBER Working Paper 20434.

Blanden, J., E. Del Bono, S. McNally, and B. Rabe. 2016. "Universal Pre-school Education: The Case of Public Funding with Private Provision." *The Economic Journal*, 126(592).

Blunt, I. 2013. "PFocus on Preventable Admissions." Nuffield Trust Quality Watch.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation."

Breivik, Anne-Lise, Emilia Del Bono, and Julie Riise. 2021. "Effects of Universal Childcare on Long-Run Health." MIMEO.

Britton, J., C. Farquharson, and L. Sibieta. 2019. "2019 Annual Report on Education Spending in England." IFS Report R162.

Bütikofer, Aline, Katrine V. Løken, and Kjell G. Salvanes. 2019. "Infant Health Care and Long-Term Outcomes." *The Review of Economics and Statistics*, 101(2): 341–354.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2020. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with a Continuous Treatment."

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." 343(6178): 1478–1485.

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).

Carneiro, Pedro, and Rita Ginja. 2016. "Partial Insurance and Investments in Children." *The Economic Journal*, 126(596): F66–F95.

Case, Anne, and Christina Paxson. 2002. "Parental Behavior And Child Health." *Health Affairs*, 21(2): 164–178.

Cattan, Sarah, Gabriella Conti, Christine Farquharson, and Rita Ginja. 2019. Institute for Fiscal Studies.

Conti, Gabriella, Giacomo Mason, and Stavros Poupakis. 2019. "Developmental Origins of Health Inequality."

Conti, Gabriella, James Heckman, and Sergio Urzua. 2010. "The education-health gradient." *American Economic Review*, 100(2): 234–38.

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

Cunha, F., J. Heckman, and S. Schennach. 2010. "Estimating the technology of cognitive and non-cognitive skill formation." *Econometrica*, 78(3): 883–931.

Cunha, F., J. Heckman, L. Lochner, and D. Masterov. 2006. "Interpreting the evidence on life cycle skill formation." In *Handbook of the Economics of Education, Volume 1.* , ed. E. Hanushek, S. Machin and L. Woessmann. Amsterdam:Elsevier Science.

Currie, Janet. 2006. "The take-up of social benefits." In *Public Policy and the Income Distribution.* , ed. D. Card A. J. Auerbach and J. M. Quigley. New York, NY:Russell Sage Foundation Publications.

Dalgaard, Carl-Johan, Casper Worm Hansen, and Holger Strulik. 2021. "Fetal origins—A life cycle model of health and aging from conception to death." *Health Economics*.

Datta Gupta, Nabanita, and Marianne Simonsen. 2010. "Non-cognitive child outcomes and universal high quality child care." *Journal of Public Economics*, 94(1): 30–43.

Davis, Katrina, Cathie Sudlow, and Matthew Hotopf. 2016. "Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses." *BMC Psychiatry*, 16.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

DfE. 2010. "Sure Start Children's Centres: Statutory Guidance." Department for Education, HMSO.

DfEE. 1999. "Sure Start: A Guide for Trailblazers." Department for Education and Employment.

DfES. 2003. "Sure Start Guidance 2004–2006: Overview and Local Delivery Arrangements." Department for Education and Skills, Nottinghamshire: DfES Publications.

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

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

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

Duncan, Greg, Ariel Kalil, Magne Mogstad, and Mari Rege. 2022. "Investing in Early Childhood Development in Preschool and at Home." *SSRN Electronic Journal*.

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." Evaluation of Children?s Centres in England (ECCE), Department for Education.

Eckenrode, John, Mary I Campa, Pamela A Morris, Charles R Henderson Jr, Kerry E Bolger, Harriet Kitzman, and David L Olds. 2017. "The prevention of child maltreatment through the nurse family partnership program: Mediating effects in a long-term follow-up study." *Child maltreatment*, 22(2): 92–99.

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

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.

Galama, Titus J, and Hans Van Kippersluis. 2019. "A theory of socio-economic disparities in health over the life cycle." *The Economic Journal*, 129(617): 338–374.

Goff, J., Hall J. Sylva K. Smith T. Smith G. Eisenstadt N. Sammons P. Evangelou M. Smees R., and K. Chu. 2013. "Evaluation of Children's Centres in England (ECCE): Strand 3: Delivery of Family Services by Children's Centres." Department for Education Report no. RR297.

González-Izquierdo, Arturo, Jenny Woodman, Lynn Copley, Jan van der Meulen, Marian Brandon, Deborah Hodes, Fiona Lecky, and Ruth Gilbert. 2010. "Variation in recording of child maltreatment in administrative records of hospital admissions for injury in England, 1997–2009." 95(11): 918–925.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*.

Gray-Lobe, Guthrie, Parag A Pathak, and Christopher R Walters. 2021. "The Long-Term Effects of Universal Preschool in Boston." National Bureau of Economic Research Working Paper 28756.

Gregg, Paul. 2008. "UK Welfare Reform 1996 to 2008 and beyond: A personalised and responsive welfare system?" The Centre for Market and Public Organisation, University of Bristol, UK The Centre for Market and Public Organisation 08/196.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *The Journal of Political Economy*, 80(2): 223–255.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review*, 103(6): 2052–86.

Henderson, Frederick W., Albert M. Collier, Wallace A. Clyde, and Floyd W. Denny. 1979. "Respiratory-Syncytial-Virus Infections, Reinfections and Immunity." *New England Journal of Medicine*, 300(10): 530–534.

Hjort, Jonas, Mikkel Sølvsten, and Miriam Wüst. 2017. "Universal Investment in Infants and Long-Run Health: Evidence from Denmark's 1937 Home Visiting Program." *American Economic Journal: Applied Economics*, 9(4).

Hoare, Peter, and Thomas Beattie. 2003. "Children with attention deficit hyperactivity disorder and attendance at hospital." *European Journal of Emergency Medicine*, 10(2): 98–100.

Hong, Kai, Kacie Dragan, and Sherry Glied. 2019. "Seeing and hearing: The impacts of New York City's universal pre-kindergarten program on the health of low-income children." *Journal of Health Economics*, 64: 93–107.

House of Commons. 2010. "Sure Start Children's Centres, Fifth Report of Session 2009–10." Children, Schools and Families Committee.

Kitzman, Harriet, David L Olds, Charles R Henderson, Carole Hanks, Robert Cole, Robert Tatelbaum, Kenneth M McConnochie, Kimberly Sidora, Dennis W Luckey, David Shaver, et al. 1997. "Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing: a randomized controlled trial." *Jama*, 278(8): 644–652.

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.

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.

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.

Meadows, Pam. 2011. "National Evaluation of Sure Start Local Programmes: An Economic Perspective." Pam Meadows and the National Evaluation of Sure Start Team, Department for Education Research Report no. DFE-RR073.

Melhuish, Edward, Jay Belsky, Alastair H Leyland, and Jacqueline Barnes. 2008. "Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: a quasi-experimental observational study." *The Lancet*, 372(9650): 1641 – 1647.

Morris, Kate, Marian Barnes, and Paul Mason. 2009. *Children, Families and Social Exclusion: New Approaches to Prevention.* Bristol:The Policy Press.

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): 1431–1437. PMID: 19542034.

NAO. 2006. "Sure Start Children's Centres." National Audit Office, HC 104 Session 2006–07, London: HMSO.

NESS. 2005. "Early impacts of Sure Start Local Programmes on Children and Families." National Evaluation of Sure Start, Department for Education and Skills.

NESS. 2008. "The impact of Sure Start Local Programmes on Three Year Olds and Their Families." National Evaluation of Sure Start, Department for Children, Schools, and Families.

NESS. 2010. "The impact of Sure Start Local Programmes on five year olds and their families." National Evaluation of Sure Start, Department for Education.

Prowse, Martin. 2008. "The Experience of Sure Start in England Draft Report prepared for EPIC by Dr Martin Prowse Social Cohesion Practical Experiences and Initiatives."

Pugh, Gillian, and Bernadette Duffy. 2010. *Contemporary Issues in the Early Years.* London:SAGE Publications Ltd.

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

Sammons, P., Hall J. Smees R., K. Smith T. Evangelou M. Eisenstadt N. Goff, J. with Sylva, and G. Smith. 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 Report no. RR495.

Smith, G., Sylva K. Sammons P. & Smith T. with Omonigho A. 2018. "Stop Start." The Sutton Trust.

Stoye, G., and B. Zaranko. 2019. "UK health spending." IFS Report R164.

Sylva, K., E. Melhuish, P. Sammons, I. Siraj-Blatchford, and B. Taggart. 2010. *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education Project (1st ed.).* London:Routledge.

Thompson, Owen. 2018. "Head Start's Long-Run Impact: Evidence from the Program's Introduction." 53(4): 1100–1139.

van den Berg, Gerard J., and Bettina M. Siflinger. 2022. "The effects of a daycare reform on health in childhood – Evidence from Sweden." *Journal of Health Economics*, 81: 102577.

Walters, Christopher R. 2015. "Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start." *American Economic Journal: Applied Economics*, 7(4).

Welshman, John. 2010. "From Head Start to Sure Start: Reflections on Policy Transfer." *Children and Society*, 24.

The Health Effects of Universal Early Childhood Interventions: Evidence from Sure Start

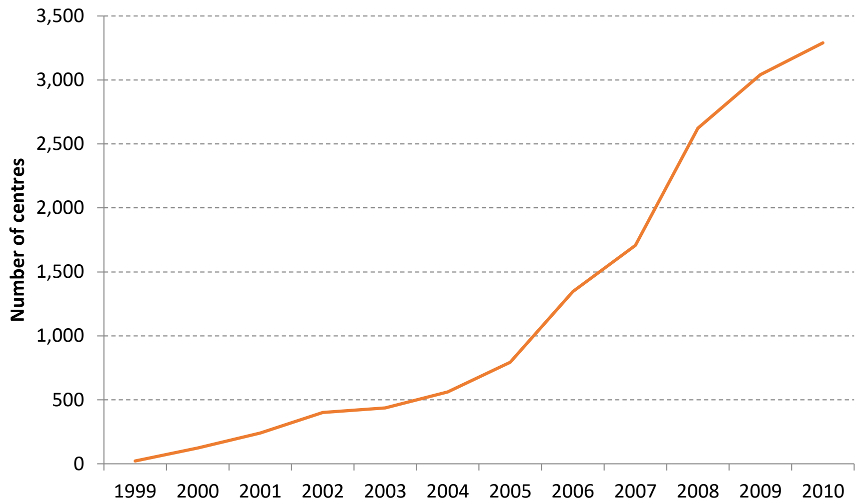
Sarah Cattan, Gabriella Conti, Christine Farquharson, Rita Ginja and Maud Pecher

October 12, 2022

MATERIAL FOR ONLINE APPENDIX

A Appendix Tables and Figures

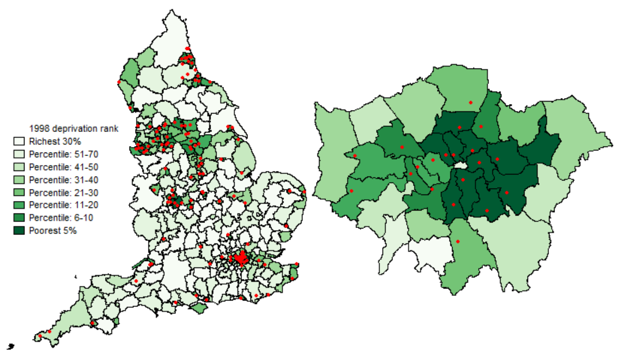
Figure A.1: Number of Sure Start centers in England



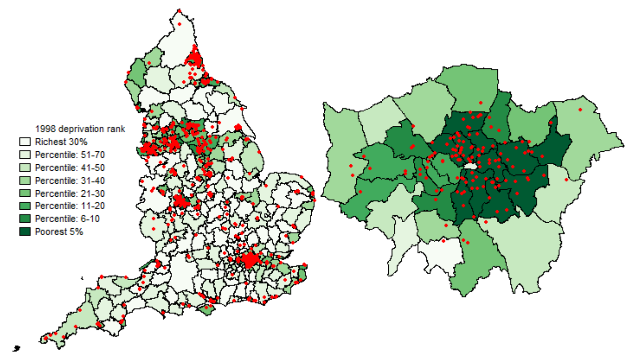
Note: The number of centers is based on centers observed in data received from the Department for Education. Since the treatment of arrangements like satellite sites was not always consistent, these numbers might not exactly match other data sources. We assume that a Sure Start Children's center (SSCC) opening at the same postcode as a Sure Start Local Programme (SSLP) replaces the SSLP; otherwise, we count both SSLPs and SSCCs between 2003 and 2006, and assume all SSLPs have closed from 2007 onward. Source: Authors' calculations using data provided by the Department for Education.

Figure A.2: Sure Start centers around England

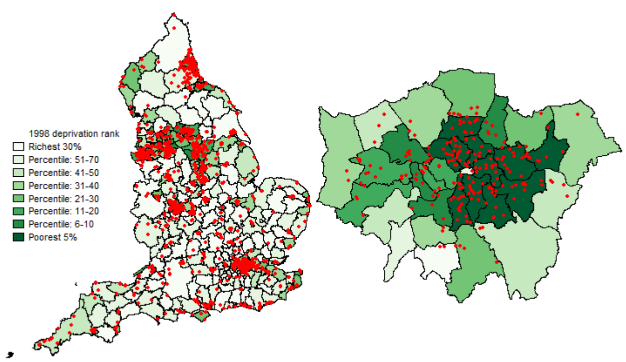
(a) SS centers in 2000



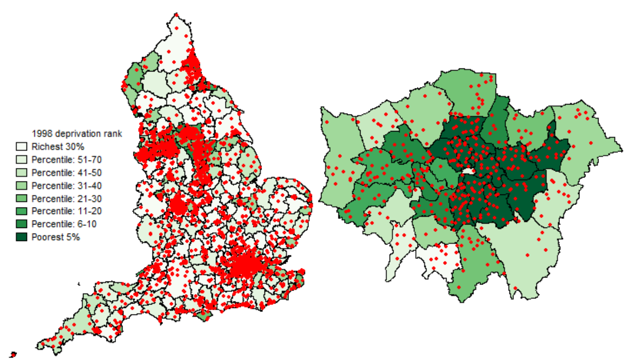
(b) SS centers in 2004



(c) SS centers in 2006

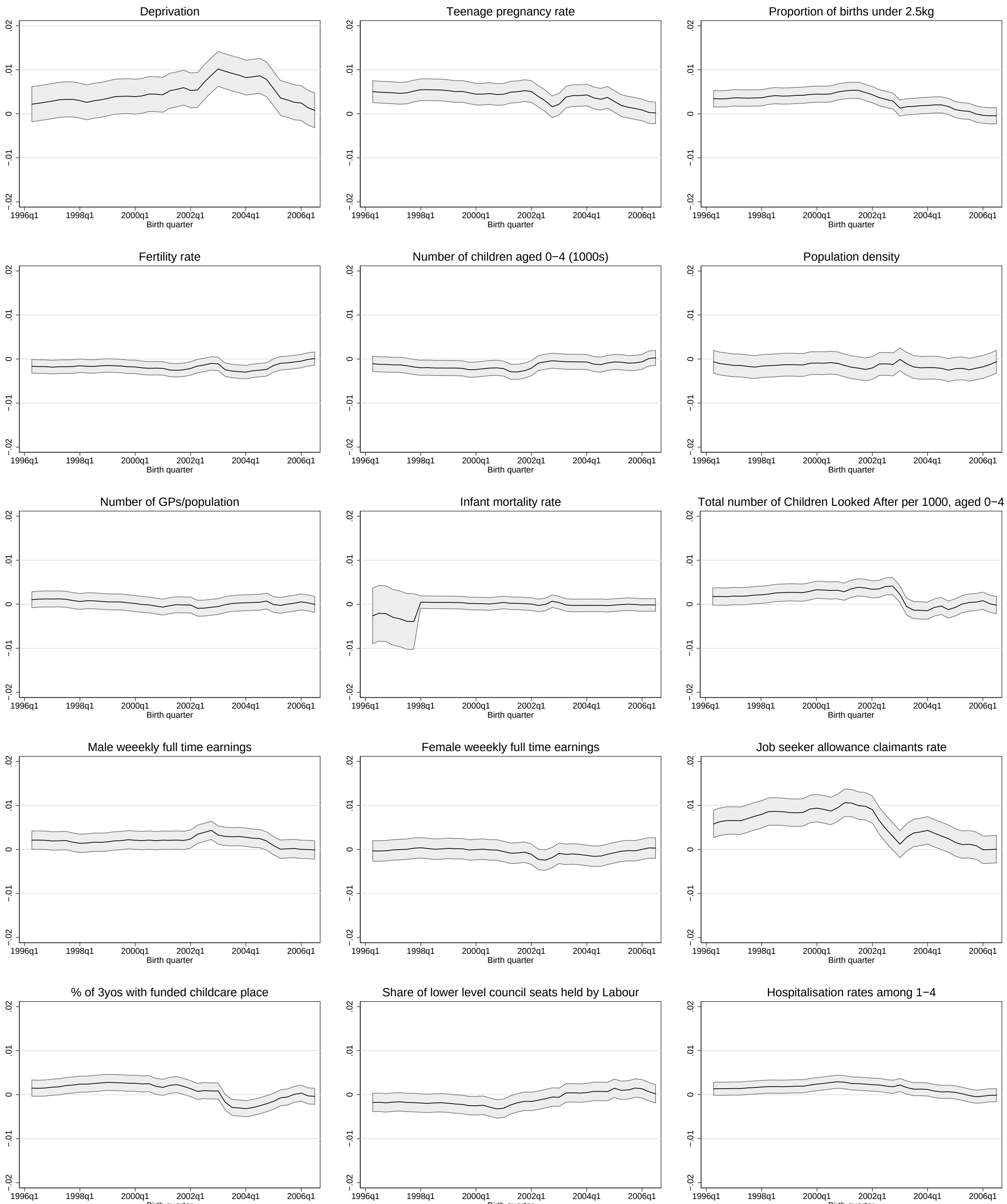


(d) SS centers in 2008



Note: Local authorities are colored by their rank in the 1998 Index of Local Deprivation, with more disadvantaged areas shaded more darkly. Each red point indicates the location of a Sure Start center (SSLP or SSCC). The maps to the right of the maps of England are zoomed-in maps of London.

Figure A.3: Regression of the change in Sure Start coverage on baseline Local Authority characteristics, 1998 - 2006



Note: These figures plot the coefficients obtained from a regression of the changes in Sure Start coverage on Local Authority specific baseline characteristics (measured in 1998) interacted with quarter-year dummies, controlling for Local Authority fixed effects. Every characteristic has been standardized to have mean 0 and standard deviation 1. The figures plot the interaction terms for each variable.

Figure A.4: Regressions of hospitalisations in pre-treated cohorts on leads of binary treatment indicator

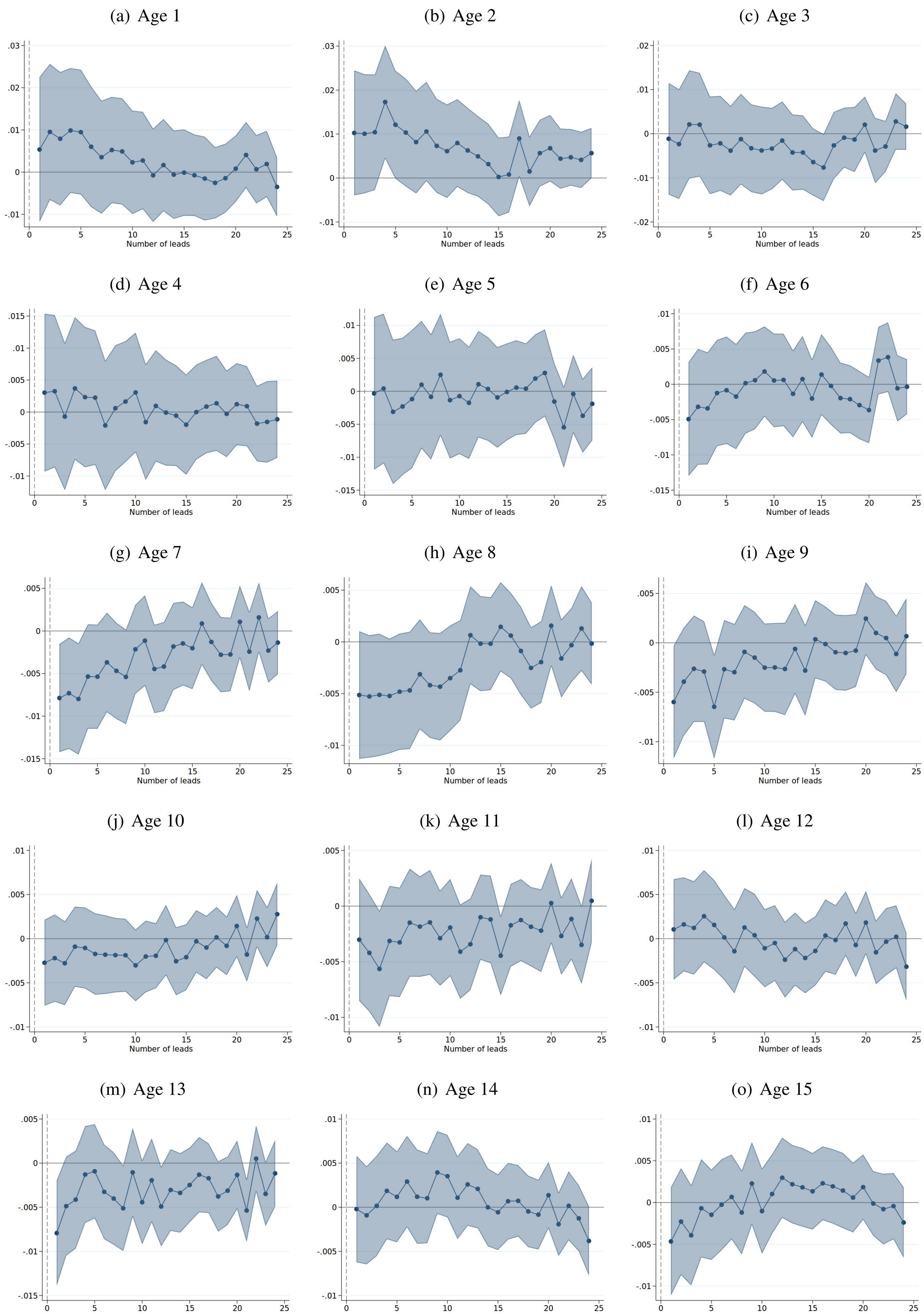


Figure A.5: Regressions of hospitalisations in pre-treated cohorts on leads of Sure Start coverage

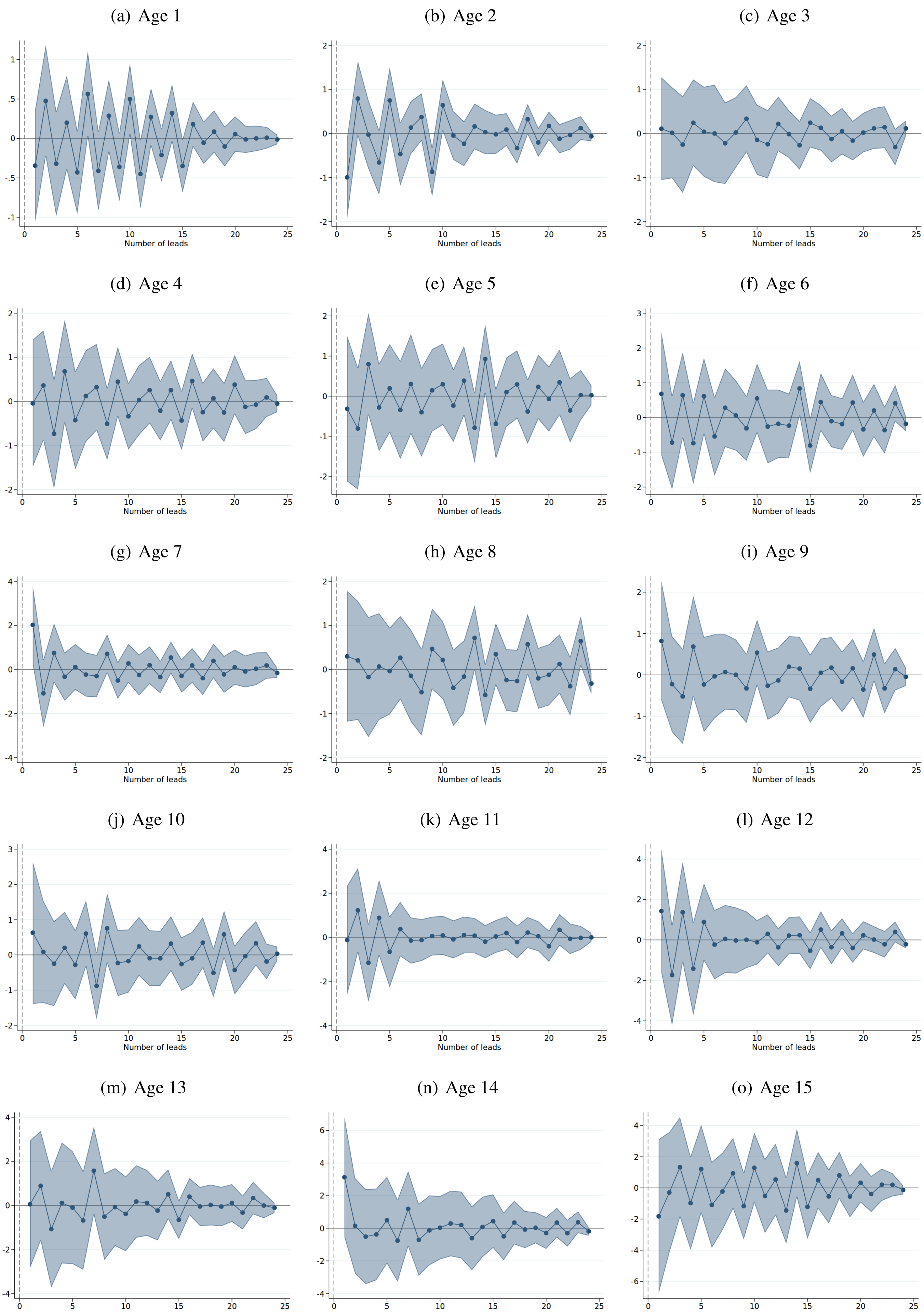
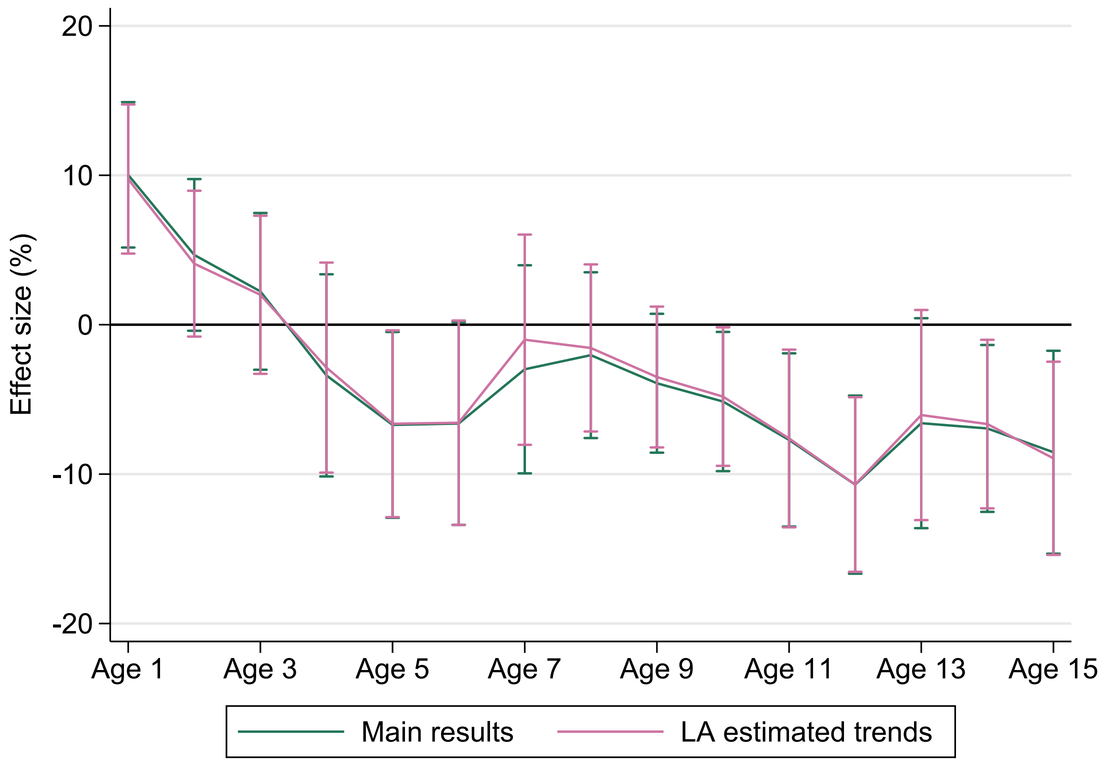
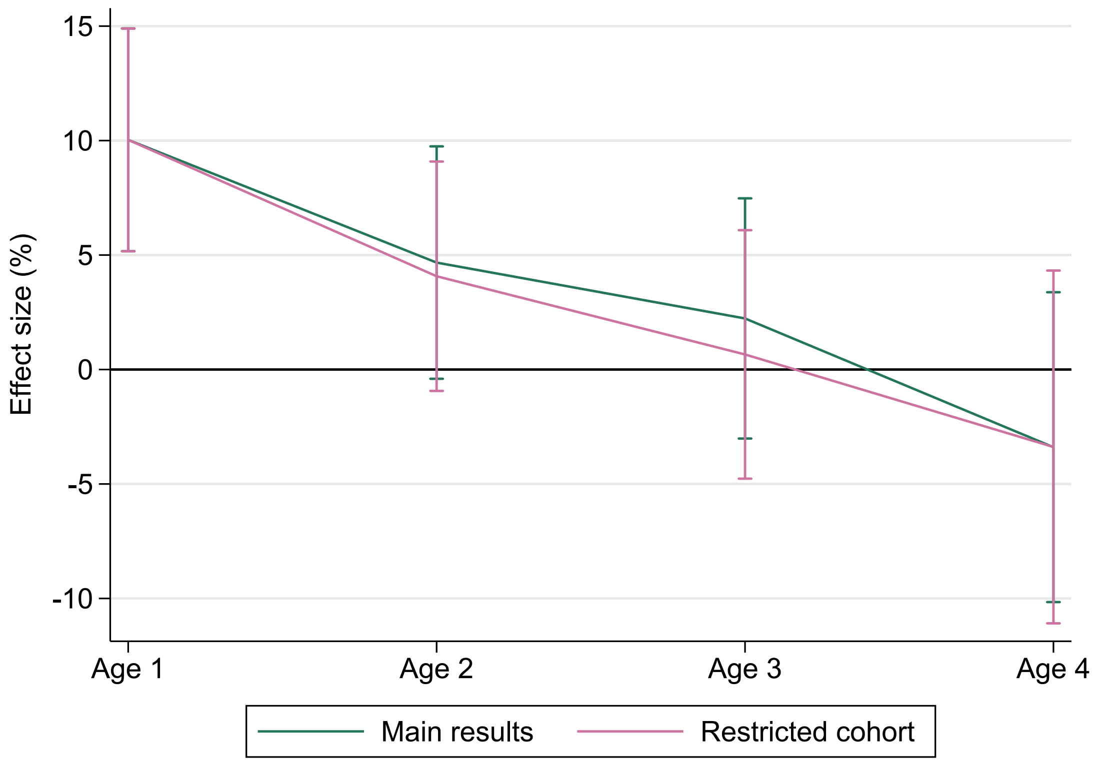


Figure A.6: Effect of an increase in Sure Start coverage on probability of any hospitalization, re-scaled by baseline probability: Baseline estimates and controlling for linear local authority trends



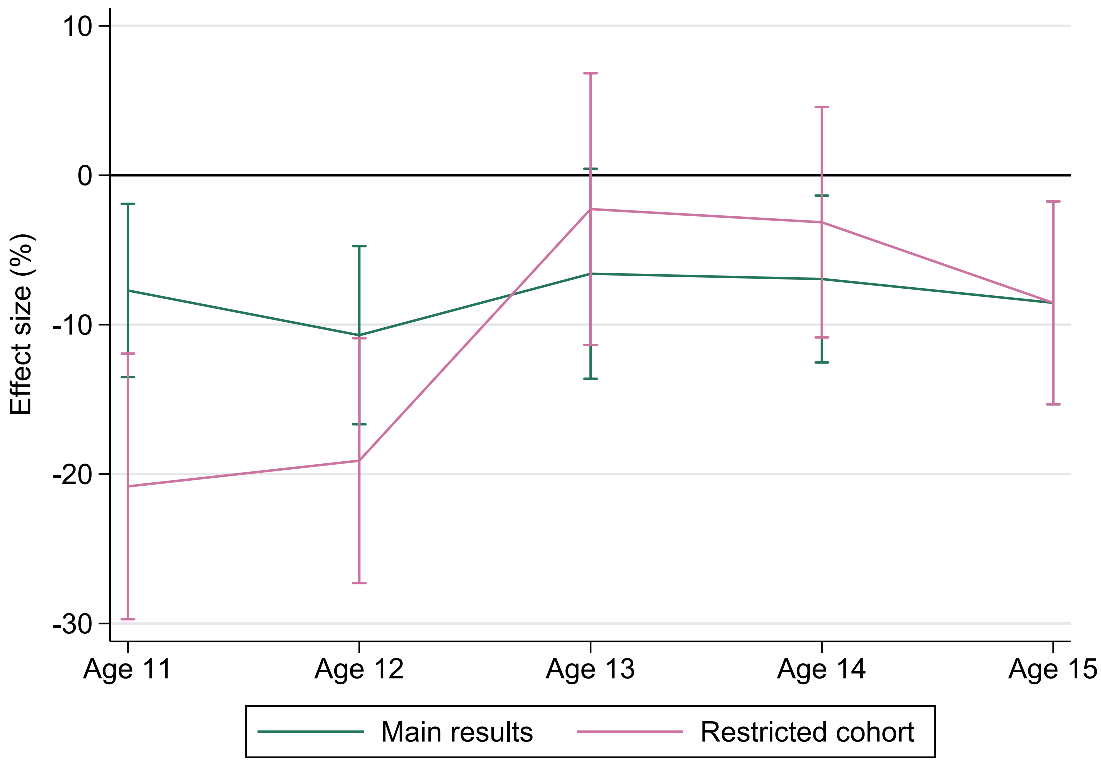
Note: Effect sizes are constructed by re-scaling the estimates by the pre-Sure Start (1996) baseline probability of a hospitalization at each age. Vertical bars indicate 90% confidence intervals. Results with LAD estimated trends additionally control for a local authority-specific linear time trend, estimated based on pre-treatment hospitalization data for each LA. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

Figure A.7: Effect of an increase in Sure Start coverage on probability of any hospitalization, re-scaled by baseline probability: Baseline estimates and estimates on a common cohort for 1- to 4-year-olds



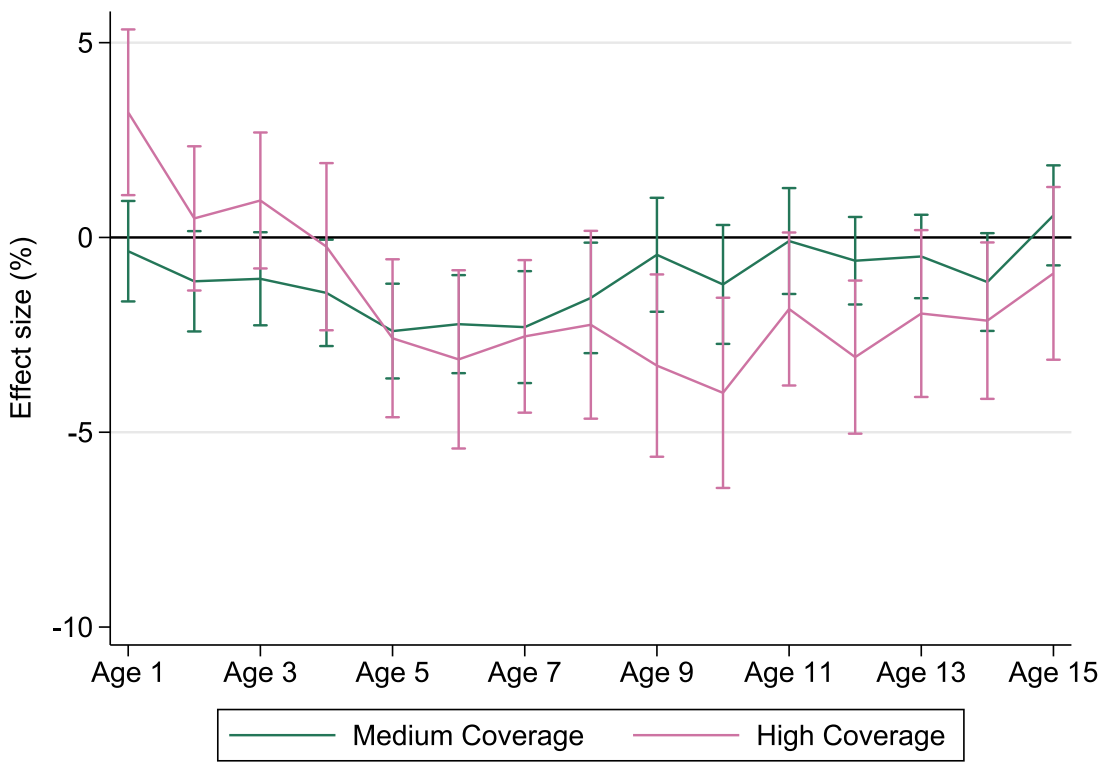
Note: The figure shows coefficients from separate regressions for each outcome age. Coefficients are re-scaled by the baseline (1996) mean for each age. Main results are estimated on cohorts as listed in Table 1. Common cohort results use a cohort of children born between April 1996 and December 2006. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

Figure A.8: Effect of an increase in Sure Start coverage on probability of any hospitalization, re-scaled by baseline probability: Baseline estimates and estimates on a common cohort for 1- to 4-year-olds



Note: The figure shows coefficients from separate regressions for each outcome age. Coefficients are re-scaled by the baseline (1996) mean for each age. Main results are estimated on cohorts as listed in Table 1. Common cohort results use a cohort of children born between January 1993 and March 2002. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

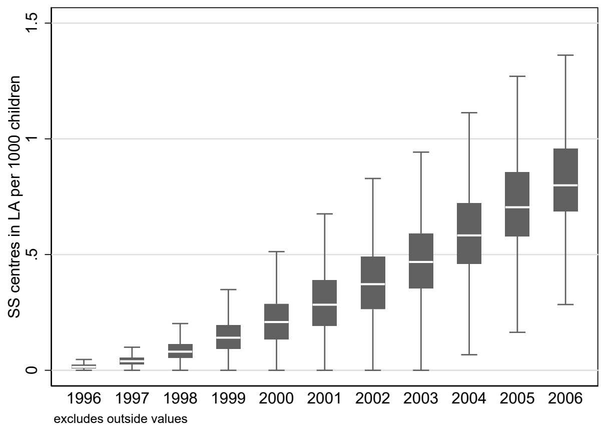
Figure A.9: Effect of an increase in Sure Start coverage on probability of any hospitalization, re-scaled by baseline probability: Non-linear estimates



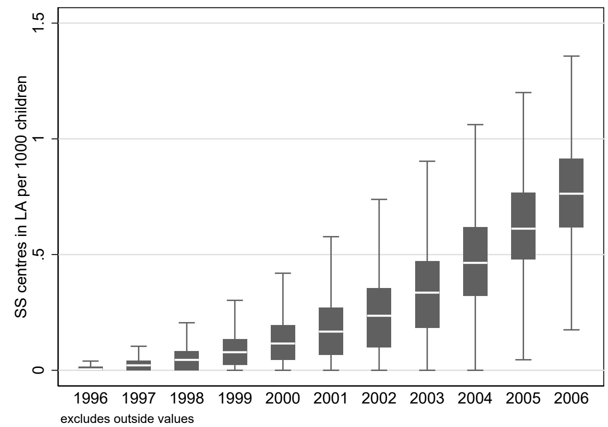
Note: The figure shows coefficients from separate regressions for each outcome age. Treatment is a pair of indicators for whether the cell experienced medium treatment (strictly positive coverage, but less than 0.25 centers per thousand children) or high treatment (more than 0.25 centers per thousand children). The omitted category is low treatment (untreated). Coefficients are re-scaled by the baseline (1996) mean for each age. Results marked with a star are significant at the 5% level. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the Department for Education's data on the rollout of Sure Start.

Figure A.10: Average coverage over the first 60 months of life, by local authority and month and year of birth: By level of deprivation

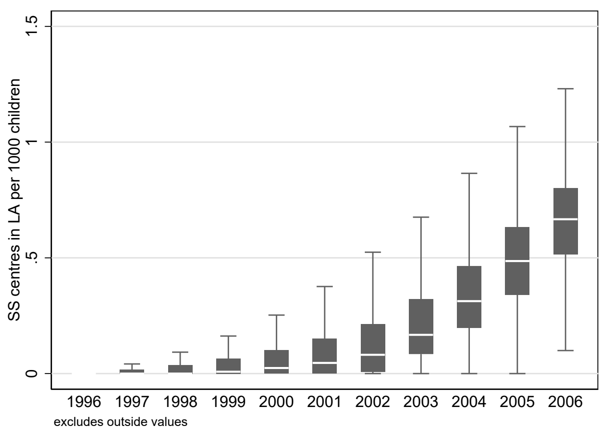
(a) Poorest 30% LSOAs



(d) Middle Deprivation Level LSOAs

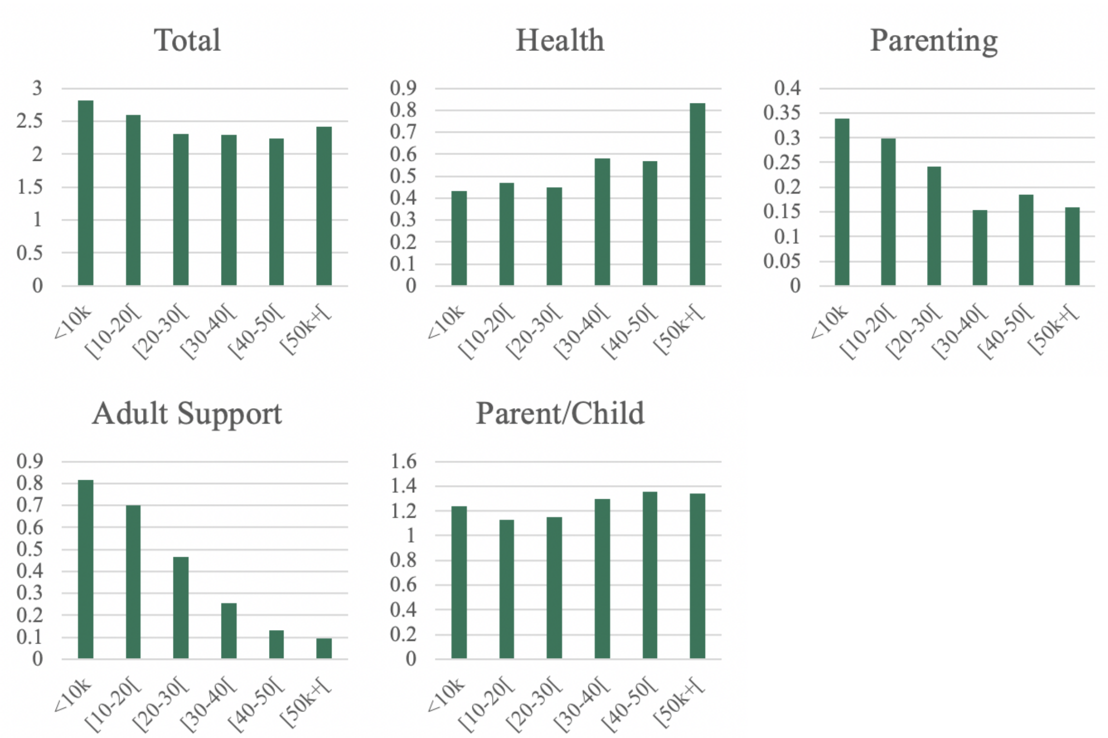


(c) Richest 30% LSOAs



Note: The figure presents the average Sure Start coverage (centers per thousand children aged 0-4 in the district) over the first five years of life for children based on their month and year of birth per LA according to the level of deprivation in the LSOA of residence in 2004. Source: Authors' calculations using data from the Department for Education and ONS population estimates.

Figure A.11: Hours spent per week at different Sure Start services by family income, 2011



Note: The figure is based on information collected by the Evaluation of Children's centers in England (ECCE) in 2011 on hours spent per week by families using different services.

Table A.1: Health benefits of early interventions

	Targeted	**Universal**
Single-component/ Specific services	- Perry Preschool Project (small-scale, Campbell et al. (2014); Conti, Heckman and Pinto (2016)): long-term health benefits	-Infant health care centres & home visiting (Bhalotra, Karlsson and Nilsson, 2017; Bütikofer, Løken and Salvanes, 2019; Hjort, Sølvsten and Wüst, 2017): long-term health benefits -Childcare (Baker, Gruber and Milligan, 2008; Breivik, Del Bono and Riise, 2021; Datta Gupta and Simonsen, 2010; Hong, Dragan and Glied, 2019; van den Berg and Siflinger, 2022): mixed health impacts
Multi-component/ Integrated services	-Abecedarian Project (small-scale, Campbell et al. (2014); Conti, Heckman and Pinto (2016)): long-term health benefits -Head-Start (large-scale, Carneiro and Ginja (2014); Frisvold and Lumeng (2011); Ludwig and Miller (2007)): long-term health benefits	**?**

Table A.2: Association between knowledge and take-up of Sure Start services and Sure Start coverage in the MCS data

	(1)	(2)	(3)	(4)
	Knows about Sure Start		Has used Sure Start	
Sure Start coverage	0.5864***	0.5156***	0.1525***	0.1268***
	(0.0383)	(0.0475)	(0.0183)	(0.0219)
Child characteristics				
Female		0.0011		0.0015
		(0.0099)		(0.0000)
Ethnicity: Asian		-0.0834***		-0.0033
		(0.0165)		(0.0000)
Ethnicity: Black		-0.0269		0.0042
		(0.0302)		(0.0000)
Ethnicity: Mixed/Other		-0.0547**		-0.0144*
		(0.0222)		(0.0000)
Family characteristics				
Partnered couple		-0.0313*		-0.0181*
		(0.0173)		(0.0000)
Mother aged 25-34 at birth		-0.0771***		-0.0436***
		(0.0141)		(0.0000)
Mother aged 35+ at birth		-0.0771***		-0.0464***
		(0.0174)		(0.0000)
Socio-economic status (index)		0.0021		-0.0116***
		(0.0055)		(0.0000)
Constant	0.1899***	0.3204***	0.0241***	0.0684***
	(0.0356)	(0.0356)	(0.0028)	(0.0028)
N	8882	8880	8882	8880
Controls and region FE?		Y		Y

Source: Millennium Cohort Study (MCS), Wave 1 (9 months old)
Note: Estimates of OLS regression of an indicator for whether a family reports knowing of Sure Start (columns 1 and 2) and using Sure Start (columns 3 and 4) on Sure Start coverage (number of Sure Start centers in the family's LA per thousand children age 0-4). In Columns 3 and 4, the regression also controls for child characteristics (gender and ethnicity dummies), family characteristics (an indicator for partnered couple, mother's age, and a socio-economic index), as well as government region fixed effects. The socio-economic index is built as: XXXX. *, ** and *** indicate significance at the 10%, 5% and 1% level respectively.

Table A.3: Predictors of take-up of Sure Start services

	(1) Any service	(2) Health services	(3) Parent-child services	(4) Parenting services	(5) Adult support
Child age (reference: age 1)					
Age 2	-0.233***	-0.261***	-0.148***	-0.032***	-0.066***
	(0.010)	(0.010)	(0.010)	(0.007)	(0.007)
Age 3	-0.355***	-0.342***	-0.242***	-0.058***	-0.112***
	(0.011)	(0.011)	(0.011)	(0.007)	(0.008)
Family SES index, interacted with:					
Child age 1	0.021***	0.016***	0.043***	-0.006	-0.014***
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)
Child age 2	-0.054***	-0.050***	-0.033***	-0.028***	-0.040***
	(0.009)	(0.009)	(0.009)	(0.006)	(0.007)
Child age 3	-0.031***	-0.025**	-0.021*	-0.017***	-0.012
	(0.010)	(0.010)	(0.011)	(0.007)	(0.007)
Child is female	-0.013	-0.005	-0.017*	-0.015***	-0.008
	(0.009)	(0.008)	(0.009)	(0.006)	(0.006)
Child ethnicity (reference: white)					
Asian	-0.068***	-0.106***	-0.064***	-0.037***	0.025**
	(0.015)	(0.014)	(0.015)	(0.009)	(0.010)
Black	0.047**	0.001	0.078***	0.046***	0.085***
	(0.018)	(0.018)	(0.019)	(0.012)	(0.013)
Mixed	-0.011	-0.097***	0.061***	0.011	-0.007
	(0.018)	(0.017)	(0.018)	(0.011)	(0.013)
Other	-0.036	-0.090***	-0.038	-0.009	-0.003
	(0.031)	(0.031)	(0.032)	(0.020)	(0.022)
Mother has partner	0.029**	0.021*	0.042***	-0.021***	-0.045***
	(0.012)	(0.012)	(0.013)	(0.008)	(0.009)
Mother age at wave 1 (reference: <25)					
25-34	0.001	0.009	-0.000	-0.004	-0.027***
	(0.012)	(0.012)	(0.012)	(0.008)	(0.009)
35 and older	-0.023	-0.032**	0.001	-0.007	-0.030***
	(0.014)	(0.014)	(0.015)	(0.009)	(0.010)
	0.752***	0.559***	0.506***	0.158***	0.238***
	(0.015)	(0.015)	(0.015)	(0.010)	(0.011)
Observations	11,835	11,835	11,835	11,835	11,835
R-squared	0.099	0.106	0.050	0.015	0.038

Note: Outcomes are indicators for whether a family reports using each service for their child at that age. The analysis pools three waves of data (age 9 months, 2 years and 3 years). The 'family SES index' is an index of socio-economic status combining income, the mother's work status and the mother's highest level of qualifications. Data on all three waves of the ECCE evaluation is used. *, ** and *** indicate significance at the 10%, 5% and 1% level respectively.

Table A.4: Expected effects of Sure Start on children's hospitalizations

Mechanism	Services	Potential effect on hospitalizations		
		Early Years (c. age 1-4)	Middle childhood (c. age 5-10)	Adolescence (c. age 11-15)
Screening and referrals to appropriate healthcare	Health services	Ambiguous effect on hospitalizations for preventable diseases	Reduction in hospitalizations for preventable diseases	
Safer home enviroment	Health services	Reduction in hospitalizations for external causes (poisoning and accidents/injuries)		
Improved emotional and behavioral development (through improved parental and formal care)	Parenting support, parent/child activities, childcare	Reduction in hospitalizations for external causes, especially accidents and injuries		
		Reduction in hospitalizations for mental health-related reasons		
Stronger immune systems	Childcare, parent/child activities	Increase in hospitalizations for infections	Reduction in hospitalizations for infections	
Increased family income and parental employment	Adult support	Reduction in hospitalizations as a result of higher family income		
		Increase in hospitalizations from less parental time at home		

Table A.5: Effect of an increase in Sure Start coverage on probability of hospitalization for any cause, weighted for number of children by LSOA

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
SS Cov	0.0254***	0.0097	0.0039	-0.0057	-0.0116*	-0.0094	-0.0033	-0.0021	-0.0047	-0.0050	-0.0091**	-0.0126***	-0.0093	-0.0109**	-0.0148**
	(0.0085)	(0.0069)	(0.0061)	(0.0074)	(0.0066)	(0.0063)	(0.0058)	(0.0043)	(0.0035)	(0.0034)	(0.0042)	(0.0047)	(0.0056)	(0.0049)	(0.0068)
Baseline mean	0.2552	0.2044	0.1791	0.1687	0.1623	0.1438	0.1260	0.1160	0.1125	0.1078	0.1089	0.1172	0.1229	0.1311	0.1410
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: The table shows coefficients from regression analysis at each outcome age weighted by the number of children in each LSOA. Observations are cells defined by the LSOA, quarter-year of birth, and sex. The model regresses an indicator for any hospitalization in a cell on Sure Start coverage, the population at the relevant age in the LSOA, an indicator for female and fixed effects for quarter–year of birth and for the LSOA of residence. Sure Start coverage SS_{qd} is defined as the number of centers per thousand children aged 0–4 in the local authority for outcomes at age 5 and older, and as the number of centers per thousand children aged 0-4 that were open in the LA d when the child was aged $a-1$ for outcomes at ages $a = 1, ..., 4$. The baseline mean (3rd row) is measured in 1996. 'Earliest' and 'latest' cohorts refer to the first and last birth cohort included in each regression. Standard errors are shown in parentheses clustered by LAD. *, ** and *** indicate significance at the 10%, 5% and 1% level.

Table A.6: Effect of an increase in Sure Start coverage on probability of hospitalization for any cause: Model controlling for trends interacted with baseline area characteristics

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
Panel A - Linear trend interacted with baseline guideline variables															
SS Cov	0.0245***	0.0074	0.0028	-0.0070	-0.0117*	-0.0103*	-0.0044	-0.0032	-0.0050	-0.0060*	-0.0083**	-0.0129***	-0.0089*	-0.0097**	-0.0124**
	(0.0074)	(0.0062)	(0.0056)	(0.0069)	(0.0061)	(0.0059)	(0.0053)	(0.0039)	(0.0032)	(0.0030)	(0.0039)	(0.0043)	(0.0053)	(0.0045)	(0.0058)
Baseline mean	0.2552	0.2044	0.1791	0.1687	0.1623	0.1438	0.1260	0.1160	0.1125	0.1078	0.1089	0.1172	0.1229	0.1311	0.1410
Panel B - Quadratic trend interacted with baseline guideline variables															
SS Cov	0.0244***	0.0073	0.0028	-0.0069	-0.0116*	-0.0102*	-0.0044	-0.0031	-0.0049	-0.0059*	-0.0083**	-0.0129***	-0.0090*	-0.0100**	-0.0129**
	(0.0074)	(0.0062)	(0.0056)	(0.0069)	(0.0061)	(0.0059)	(0.0053)	(0.0039)	(0.0032)	(0.0031)	(0.0039)	(0.0043)	(0.0053)	(0.0045)	(0.0058)
Baseline mean	0.2552	0.2044	0.1791	0.1687	0.1623	0.1438	0.1260	0.1160	0.1125	0.1078	0.1089	0.1172	0.1229	0.1311	0.1410
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: The table shows coefficients from regression analysis at each outcome age. Observations are cells defined by the LSOA, quarter-year of birth, and sex. The model regresses an indicator for any hospitalization in a cell on Sure Start coverage, the population at the relevant age in the LSOA, an indicator for female and fixed effects for quarter–year of birth and for the LSOA of residence, as well as interactions between three rollout guideline variables measured in 1998 and a linear trend (Panel A) or a quadratic trend (Panel B). The three rollout guidelines variables are: area deprivation level, proportion of births with low birth weight, and fertility rate of teenage mothers living in the area. Sure Start coverage SS_{qd} is defined as the number of centers per thousand children aged 0–4 in the local authority for outcomes at age 5 and older, and as the number of centers per thousand children aged 0-4 that were open in the LA d when the child was aged $a-1$ for outcomes at ages $a = 1, ..., 4$. The baseline mean (3rd row) is measured in 1996. 'Earliest' and 'latest' cohorts refer to the first and last birth cohort included in each regression. Standard errors are shown in parentheses clustered by LAD. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; $+$, $++$ and $+++$ indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure described in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

Table A.7: Effect of an increase in Sure Start coverage on probability of hospitalization for congenital chromosomal defects between 2 and 11 months

	(1) Coverage at birth	(2) Avg. coverage ages 0-4
SS coverage	0.0011	0.0006
	(0.0011)	(0.0010)
N	2,625,280	2,625,280
Baseline mean	0.0237	0.0237
Earliest cohort	Apr. 1997	Apr. 1997
Latest cohort	Dec. 2006	Dec. 2006

Note: See notes to Table 1. The first column defines Sure Start treatment based on the number of centers per thousand children in the LA at the time of the child's birth. The second column uses the average coverage over the first five years of life, as we use in our main results (note that this means some treatment postdates the outcome, which is measured between 2 and 11 months). *, ** and *** indicate significance at the 10%, 5% and 1% level respectively.

Table A.8: Effect of an increase in Sure Start coverage on number of hospitalizations for any cause and length of stay for such hospitalizations

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
A - Dependent variable: Number of hospitalisations at the cell level															
SS Cov	0.0549***	0.0260*	0.0093	-0.0056	-0.0202*	-0.0126	-0.0027	0.0011	-0.0033	-0.0001	-0.0024	-0.0053	-0.0029	0.0025	-0.0102
	(0.0189)	(0.0134)	(0.0120)	(0.0128)	(0.0122)	(0.0115)	(0.0104)	(0.0082)	(0.0083)	(0.0078)	(0.0088)	(0.0093)	(0.0112)	(0.0104)	(0.0151)
Baseline mean	0.4306	0.3224	0.2791	0.2556	0.2420	0.2108	0.1846	0.1755	0.1708	0.1651	0.1709	0.1840	0.1950	0.2125	0.2335
B - Dependent variable: Length of stay of hospital admissions															
SS Cov	0.0110	0.0120	0.0128	-0.0074	-0.0125	-0.0025	-0.0046	-0.0028	-0.0053	-0.0101***	-0.0126***	-0.0062	-0.0008	-0.0178	-0.0265
	(0.0121)	(0.0139)	(0.0102)	(0.0077)	(0.0086)	(0.0063)	(0.0053)	(0.0048)	(0.0048)	(0.0035)	(0.0034)	(0.0047)	(0.0064)	(0.0111)	(0.0241)
Baseline mean	0.1631	0.1102	0.0959	0.0760	0.0603	0.0570	0.0459	0.0451	0.0397	0.0367	0.0399	0.0485	0.0547	0.0634	0.0772
N	2822176	3084704	3347232	3609760	3675392	3675392	3675392	3675392	3675392	3675392	3478496	3215968	2953440	2690912	2428384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: The table shows coefficients from regression analysis at each outcome age. Observations are cells defined by the LSOA, quarter-year of birth, and sex. The model regresses the number of hospitalizations (Panel A) and length of hospital (Panel B) in a cell on Sure Start coverage, the population at the relevant age in the LSOA, an indicator for female and fixed effects for quarter–year of birth and for the LSOA of residence. Both outcome variables include 0s for no hospitalization. Sure Start coverage SS_{qd} is defined as the number of centers per thousand children aged 0–4 in the local authority for outcomes at age 5 and older, and as the number of centers per thousand children aged 0-4 that were open in the LA d when the child was aged $a-1$ for outcomes at ages $a = 1, ..., 4$. The baseline mean (3rd row) is measured in 1996. 'Earliest' and 'latest' cohorts refer to the first and last birth cohort included in each regression. Standard errors are shown in parentheses clustered by LAD. *, ** and *** indicate significance at the 10%, 5% and 1% level.

Table A.9: Proportion and sum of negative weights used in the calculation of the treatment effects in the Two Way Fixed Effects model

	Number of weights	Proportion negative weights	Sum of positive weights	Sum of negative weights
Age 1	729,253	51%	1.174	-0.174
Age 2	859,654	53%	1.241	-0.241
Age 3	990,458	55%	1.325	-0.325
Age 4	1,121,686	55%	1.416	-0.416
Age 5	1,252,950	55%	1.532	-0.532
Age 6	1,252,950	55%	1.532	-0.532
Age 7	1,252,950	55%	1.532	-0.532
Age 8	1,252,950	55%	1.532	-0.532
Age 9	1,252,950	55%	1.532	-0.532
Age 10	1,252,950	55%	1.532	-0.532
Age 11	1,154,502	56%	1.445	-0.445
Age 12	1,023,238	57%	1.331	-0.331
Age 13	892,333	56%	1.235	-0.235
Age 14	761,617	55%	1.162	-0.162
Age 15	639,741	54%	1.134	-0.134

Table A.10: Comparison of binary treatment effect estimates using the TWFE and Borusyak, Jaravel and Spiess (2021) estimators

Age of admission	Estimator	$1(SS_{dq} > 0))$		$1(SS_{dq} > 0.1)$		$1(SS_{dq} > 0.25)$	
Age 1	TWFE	-0.001	(0.002)	0.005**	(0.002)	0.009***	(0.003)
	BJS	0.006**	(0.003)	0.008***	(0.003)	0.011***	(0.003)
Age 2	TWFE	-0.002	(0.002)	0.000	(0.002)	0.003*	(0.002)
	BJS	0.006**	(0.003)	0.005*	(0.002)	0.005**	(0.002)
Age 3	TWFE	-0.002*	(0.001)	0.001	(0.001)	0.004***	(0.001)
	BJS	0.006**	(0.002)	0.000	(0.003)	0.004	(0.002)
Age 4	TWFE	-0.003*	(0.001)	0.001	(0.001)	0.002	(0.002)
	BJS	0.002	(0.002)	0.003	(0.002)	0.002	(0.003)
Age 5	TWFE	-0.004***	(0.001)	0.001	(0.001)	0.000	(0.002)
	BJS	0.002	(0.002)	0.004**	(0.002)	-0.002	(0.002)
Age 6	TWFE	-0.003***	(0.001)	0.001	(0.001)	-0.001	(0.002)
	BJS	0.000	(0.002)	0.001	(0.003)	0.001	(0.003)
Age 7	TWFE	-0.003**	(0.001)	0.002	(0.001)	0.000	(0.001)
	BJS	0.001	(0.002)	0.003	(0.003)	0.002	(0.004)
Age 8	TWFE	-0.002	(0.001)	0.002	(0.001)	-0.001	(0.001)
	BJS	-0.001	(0.002)	0.001	(0.002)	-0.001	(0.002)
Age 9	TWFE	-0.000	(0.001)	-0.000	(0.001)	-0.003**	(0.001)
	BJS	0.004**	(0.002)	-0.001	(0.001)	-0.003**	(0.001)
Age 10	TWFE	-0.001	(0.001)	-0.001*	(0.001)	-0.003**	(0.001)
	BJS	0.000	(0.003)	-0.002	(0.001)	-0.002	(0.002)
Age 11	TWFE	0.000	(0.001)	-0.003***	(0.001)	-0.002	(0.001)
	BJS	0.002	(0.002)	0.002	(0.002)	-0.001	(0.001)
Age 12	TWFE	-0.000	(0.001)	-0.004***	(0.001)	-0.003**	(0.001)
	BJS	0.000	(0.003)	-0.003*	(0.001)	-0.003**	(0.001)
Age 13	TWFE	-0.000	(0.001)	-0.001	(0.001)	-0.002	(0.001)
	BJS	-0.004**	(0.002)	-0.002	(0.001)	-0.002*	(0.001)
Age 14	TWFE	-0.001	(0.001)	-0.001	(0.001)	-0.001	(0.001)
	BJS	-0.003**	(0.001)	-0.002	(0.001)	-0.002	(0.001)
Age 15	TWFE	0.001	(0.001)	-0.001	(0.001)	-0.002	(0.001)
	BJS	-0.001	(0.001)	-0.003**	(0.001)	-0.003**	(0.001)

Note: This table reports the coefficients associated with a binary measure of Sure Start coverage estimated in the TWFE model and using the Borusyak, Jaravel and Spiess (2021) estimator. We consider three different definitions of this binary measure of Sure Start coverage: an indicator for whether SS_{dq} is above 0 (results reported in column 3 of the table), an indicator for whether SS_{dq} is above 0.1 (column 4) and an indicator for whether it is above 0.25 (column 5). With both estimators, we control for a gender dummy and the number of individuals of age a when the dependent variable measures hospitalizations at age a. The TWFE model also controls for neighborhood (defined at the LSOA level) and cohort (defined as the year-quarter of birth) level. *, ** and *** indicate significance at the 10%, 5% and 1% level respectively.

Table A.11: Effect of an increase in Sure Start coverage on probability of hospitalization due to specific conditions

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
Panel A: Any hospitalisation for infectious illnesses															
SS Cov	0.0202***	0.0074	0.0015	-0.0035	-0.0071**	-0.0067***	-0.0017	-0.0016	-0.0006	-0.0005	-0.0016	-0.0020	-0.0007	-0.0011	-0.0015
	(0.0071)++	(0.0048)	(0.0037)	(0.0039)	(0.0029)++	(0.0021)++	(0.0015)	(0.0011)	(0.0012)	(0.0013)	(0.0012)	(0.0014)	(0.0015)	(0.0015)	(0.0017)
Baseline mean	0.1208	0.0808	0.0635	0.0494	0.0466	0.0361	0.0275	0.0231	0.0225	0.0187	0.0160	0.0159	0.0145	0.0136	0.0147
Panel B: Any hospitalisation for ACS related cause															
SS Cov	0.0123**	0.0072**	0.0038*	0.0021	0.0007	-0.0001	-0.0003	-0.0003	-0.0015**	-0.0013*	-0.0016*	-0.0016	-0.0008	-0.0005	-0.0013
	(0.0051)++	(0.0031)+	(0.0022)	(0.0017)	(0.0013)	(0.0010)	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0010)	(0.0011)	(0.0009)	(0.0012)
Baseline mean	0.0609	0.0375	0.0235	0.0162	0.0129	0.0101	0.0089	0.0093	0.0092	0.0086	0.0076	0.0075	0.0072	0.0071	0.0079
Panel C: Any hospitalisation for an external cause															
SS Cov	-0.0041**	-0.0052***	-0.0057***	-0.0062***	-0.0056***	-0.0056***	-0.0040***	-0.0024**	-0.0017*	-0.0036***	-0.0023*	-0.0018	-0.0001	-0.0007	-0.0029
	(0.0019)+	(0.0018)+++	(0.0016)+++	(0.0015)+++	(0.0013)+++	(0.0012)+++	(0.0010)+++	(0.0010)++	(0.0010)	(0.0010)+++	(0.0013)	(0.0014)	(0.0017)	(0.0022)	(0.0029)
Baseline mean	0.0397	0.0395	0.0305	0.0255	0.0257	0.0245	0.0227	0.0214	0.0207	0.0218	0.0231	0.0260	0.0291	0.0335	0.0357
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: See notes to Table 1. Cause-specific results are based on the primary diagnosis at the time of admission. External admissions include ICD-10 codes in groups S, T, V and Y. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; +, ++ and +++ indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

Table A.12: Effect of an increase in Sure Start coverage on probability of hospitalization due to external causes

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
Panel A: Any hospitalisation for poisoning															
SS Cov	-0.0033***	-0.0025**	-0.0024**	-0.0008	-0.0002	-0.0001	-0.0004	0.0003	0.0006*	-0.0003	0.0003	0.0006	0.0004	0.0015	-0.0009
	(0.0011)+++	(0.0012)++	(0.0010)++	(0.0007)	(0.0006)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0005)	(0.0009)	(0.0014)	(0.0019)
			2.4												
Baseline mean	0.0187	0.0188	0.0114	0.0070	0.0058	0.0046	0.0039	0.0034	0.0031	0.0028	0.0031	0.0041	0.0065	0.0110	0.0139
Panel B: Any hospitalisation for injuries															
SS Cov	-0.0010	-0.0030**	-0.0033***	-0.0056***	-0.0054***	-0.0055***	-0.0035***	-0.0026***	-0.0022***	-0.0033***	-0.0026**	-0.0024*	-0.0006	-0.0018	-0.0027
	(0.0012)	(0.0012)++	(0.0010)+++	(0.0013)+++	(0.0011)+++	(0.0010)+++	(0.0009)+++	(0.0009)+++	(0.0009)+++	(0.0010)+++	(0.0013)	(0.0014)	(0.0014)	(0.0018)	(0.0022)
Baseline mean	0.0216	0.0214	0.0194	0.0187	0.0201	0.0200	0.0189	0.0181	0.0178	0.0191	0.0202	0.0221	0.0230	0.0232	0.0227
Panel C: Any hospitalisation for fractures															
SS Cov	-0.0007*	-0.0009**	-0.0007*	-0.0008	-0.0016***	-0.0028***	-0.0009*	-0.0012**	-0.0013***	-0.0013***	-0.0018***	-0.0010	0.0005	0.0008	0.0005
	(0.0004)	(0.0004)	(0.0004)	(0.0006)	(0.0005)+++	(0.0006)+++	(0.0005)	(0.0005)+	(0.0005)+++	(0.0005)+++	(0.0007)+++	(0.0007)	(0.0009)	(0.0010)	(0.0013)
Baseline mean	0.0054	0.0065	0.0068	0.0080	0.0101	0.0111	0.0108	0.0100	0.0102	0.0115	0.0121	0.0130	0.0138	0.0134	0.0119
Panel D: Any hospitalisation for head injuries															
SS Cov	-0.0007	-0.0016*	-0.0010	-0.0028***	-0.0019***	-0.0005	-0.0006*	-0.0007	-0.0006**	-0.0006*	0.0000	-0.0005	-0.0004	-0.0011	-0.0017*
	(0.0008)	(0.0008)	(0.0006)	(0.0007)+++	(0.0006)+++	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0005)	(0.0005)	(0.0007)	(0.0009)
Baseline mean	0.0084	0.0079	0.0066	0.0052	0.0046	0.0037	0.0031	0.0029	0.0028	0.0029	0.0031	0.0038	0.0044	0.0049	0.0051
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: See notes to Table 1. Cause-specific results are based on the primary diagnosis at the time of admission. External admissions include ICD-10 codes in groups S, T, V and Y. Poisonings include ICD-10 codes T15-T98; injuries include codes S00-T14; fractures include codes S02, S12, S22, S32, S42, S52, S62, S72, S82, S92, T02, T10, T14; and head injuries include codes S00-S09. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; +, ++ and +++ indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

Table A.13: Effect of an increase in Sure Start coverage on probability of hospitalization for mental health

	(1) Age 11	(2) Age 12	(3) Age 13	(4) Age 14	(5) Age 15
SS Cov	-0.0003	-0.0007**	-0.0016***	-0.0019**	-0.0010
	(0.0002)	(0.0003)++	(0.0005)+++	(0.0009)++	(0.0013)
Baseline mean	0.0007	0.0013	0.0026	0.0042	0.0049
N	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: See notes to Table 1. Cause-specific results are based on the primary diagnosis at the time of admission. Mental health admissions are based on ICD-10 group F. Results for younger ages are omitted because of very low prevalence. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; +, ++ and + + + indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

Table A.14: Effect of an increase in Sure Start coverage on probability of hospitalization: By admission route

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
Panel A: Emergency admissions															
SS Cov	0.0276***	0.0132**	0.0063	-0.0029	-0.0060	-0.0065*	-0.0042	-0.0037	-0.0056**	-0.0057**	-0.0064*	-0.0101***	-0.0045	-0.0069*	-0.0079
	(0.0082)+++	(0.0066)	(0.0056)	(0.0057)	(0.0042)	(0.0037)	(0.0028)	(0.0026)	(0.0025)	(0.0028)	(0.0034)	(0.0038)++	(0.0043)	(0.0040)	(0.0052)
Baseline mean	0.2118	0.1531	0.1132	0.0897	0.0789	0.0677	0.0623	0.0627	0.0646	0.0639	0.0639	0.0681	0.0692	0.0753	0.0817
Panel B: Elective admissions															
SS Cov	0.0019	-0.0039	-0.0019	-0.0041	-0.0055	-0.0049	0.0001	0.0007	-0.0009	-0.0009	-0.0033	-0.0044*	-0.0042	-0.0018	-0.0039
	(0.0031)	(0.0025)	(0.0029)	(0.0040)	(0.0048)	(0.0048)	(0.0046)	(0.0032)	(0.0026)	(0.0020)	(0.0021)	(0.0023)	(0.0032)	(0.0035)	(0.0038)
Baseline mean	0.0741	0.0753	0.0875	0.0982	0.1014	0.0909	0.0756	0.0641	0.0583	0.0539	0.0556	0.0606	0.0664	0.0697	0.0756
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: See notes to Table 1. Emergency admission routes include emergency room services or emergency referrals from GPs; elective admissions include pre-planned inpatient care. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; +, ++ and + + + indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

Table A.15: Effect of an increase in Sure Start coverage on probability of hospitalization for any cause, by gender

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
SS Cov: Boys δ_B	0.0291***	0.0102	0.0034	-0.0105	-0.0138**	-0.0126**	-0.0053	-0.0053	-0.0062*	-0.0060*	-0.0092**	-0.0153***	-0.0158***	-0.0212***	-0.0249***
	(0.0081)+++	(0.0066)	(0.0058)	(0.0073)	(0.0063)+	(0.0061)	(0.0054)	(0.0039)	(0.0033)	(0.0031)	(0.0041)++	(0.0046)+++	(0.0058)++	(0.0053)+++	(0.0065)+++
SS Cov: Girls δ_G	0.0221***	0.0089	0.0046	-0.0010	-0.0080	-0.0064	-0.0022	0.0006	-0.0026	-0.0051	-0.0077**	-0.0098**	-0.0004	0.0030	0.0009
	(0.0072)+++	(0.0062)	(0.0058)	(0.0067)	(0.0061)	(0.0059)	(0.0053)	(0.0041)	(0.0032)	(0.0031)	(0.0038)	(0.0041)++	(0.0050)	(0.0043)	(0.0067)
p-values:															
$H_A : \delta_B \neq \delta_G$	0.016	0.616	0.588	0.000	0.000	0.000	0.024	0.000	0.003	0.466	0.311	0.010	0.000	0.000	0.000
H_A : Diff. effect size	0.798	0.592	0.384	0.004	0.253	0.106	0.247	0.003	0.062	0.717	0.746	0.184	0.000	0.000	0.000
Baseline mean:															
Boys	0.2863	0.2348	0.2099	0.1982	0.1871	0.1652	0.1419	0.1302	0.1255	0.1223	0.1220	0.1300	0.1283	0.1292	0.1320
Girls	0.2241	0.1739	0.1482	0.1392	0.1375	0.1225	0.1101	0.1018	0.0995	0.0933	0.0959	0.1044	0.1174	0.1331	0.1499
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: See notes to Table 1. Each regression interacts Sure Start coverage with indicators for whether the cell contains boys or girls (coverage on its own is not included in this model.) 'Difference p-value' tests the equality of the coefficients for coverage interacted with boys and with girls. 'Effect size difference p-value' tests the equality of the effect size (coefficients weighted by subgroup baseline mean) for coverage interacted with boys and with girls. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; $+$, $++$ and $+++$ indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

Table A.16: Effect of an increase in Sure Start coverage on probability of hospitalization for specific conditions, by gender

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
Panel A: Any hospitalisation for infectious illnesses															
SS Cov: Boys δ_B	0.0255***	0.0102**	0.0052	-0.0020	-0.0056*	-0.0058**	-0.0002	-0.0009	0.0003	0.0002	-0.0005	-0.0005	-0.0003	0.0003	-0.0004
	(0.0078)	(0.0052)	(0.0040)	(0.0043)	(0.0030)	(0.0023)	(0.0015)	(0.0011)	(0.0013)	(0.0013)	(0.0013)	(0.0016)	(0.0016)	(0.0017)	(0.0019)
SS Cov: Girls δ_G	0.0150**	0.0045	-0.0022	-0.0049	-0.0087***	-0.0077***	-0.0032**	-0.0023**	-0.0015	-0.0012	-0.0026**	-0.0035***	-0.0011	-0.0026	-0.0027
	(0.0065)	(0.0046)	(0.0035)	(0.0036)	(0.0028)	(0.0021)	(0.0015)	(0.0011)	(0.0012)	(0.0013)	(0.0012)	(0.0013)	(0.0016)	(0.0016)	(0.0019)
p-values:															
$H_A : \delta_B \neq \delta_G$	0.000	0.001	0.000	0.016	0.000	0.011	0.000	0.015	0.002	0.015	0.001	0.000	0.383	0.034	0.165
H_A : Diff. effect size	0.086	0.069	0.000	0.002	0.000	0.000	0.000	0.007	0.001	0.013	0.001	0.000	0.425	0.062	0.299
Baseline mean:															
Boys	0.1387	0.0944	0.0747	0.0585	0.0538	0.0404	0.0289	0.0245	0.0232	0.0194	0.0161	0.0155	0.0138	0.0121	0.0123
Girls	0.1029	0.0672	0.0522	0.0403	0.0394	0.0319	0.0260	0.0216	0.0218	0.0181	0.0160	0.0163	0.0152	0.0152	0.0172
Panel B: Any hospitalisation for ACS related cause															
SS Cov: Boys δ_B	0.0138**	0.0086***	0.0049**	0.0030*	0.0014	0.0003	0.0001	-0.0000	-0.0013	-0.0010	-0.0016*	-0.0015	-0.0009	-0.0012	-0.0028**
	(0.0055)	(0.0033)	(0.0023)	(0.0018)	(0.0014)	(0.0011)	(0.0008)	(0.0007)	(0.0008)	(0.0008)	(0.0009)	(0.0011)	(0.0012)	(0.0010)	(0.0013)
SS Cov: Girls δ_G	0.0107**	0.0058*	0.0027	0.0012	0.0000	-0.0005	-0.0008	-0.0007	-0.0018**	-0.0016**	-0.0015*	-0.0017*	-0.0006	0.0003	0.0001
	(0.0049)	(0.0031)	(0.0021)	(0.0016)	(0.0013)	(0.0009)	(0.0007)	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0010)	(0.0012)	(0.0010)	(0.0014)
p-values:															
$H_A : \delta_B \neq \delta_G$	0.106	0.025	0.007	0.002	0.003	0.128	0.025	0.109	0.176	0.076	0.785	0.741	0.557	0.061	0.008
H_A : Diff. effect size	0.871	0.681	0.277	0.045	0.024	0.119	0.016	0.087	0.042	0.010	0.578	0.368	0.576	0.052	0.003
Baseline mean:															
Boys	0.0679	0.0432	0.0271	0.0188	0.0148	0.0115	0.0099	0.0102	0.0099	0.0095	0.0083	0.0082	0.0073	0.0066	0.0069
Girls	0.0538	0.0318	0.0198	0.0137	0.0110	0.0088	0.0080	0.0085	0.0085	0.0076	0.0068	0.0069	0.0071	0.0075	0.0089
Panel C: Any hospitalisation for External related cause															
SS Cov: Boys δ_B	-0.0029	-0.0039**	-0.0044**	-0.0064***	-0.0059***	-0.0060***	-0.0039***	-0.0037***	-0.0031***	-0.0042***	-0.0046***	-0.0060***	-0.0097***	-0.0135***	-0.0204***
	(0.0021)	(0.0020)	(0.0018)	(0.0016)	(0.0014)	(0.0012)	(0.0011)	(0.0010)	(0.0010)	(0.0011)	(0.0014)	(0.0017)	(0.0019)	(0.0026)	(0.0033)
SS Cov: Girls δ_G	-0.0052***	-0.0066***	-0.0070***	-0.0061***	-0.0053***	-0.0052***	-0.0041***	-0.0012	-0.0004	-0.0030***	-0.0001	0.0024*	0.0094***	0.0121***	0.0145***
	(0.0018)	(0.0018)	(0.0016)	(0.0015)	(0.0013)	(0.0012)	(0.0010)	(0.0011)	(0.0010)	(0.0010)	(0.0013)	(0.0014)	(0.0018)	(0.0025)	(0.0036)
p-values:															
$H_A : \delta_B \neq \delta_G$	0.049	0.007	0.002	0.585	0.252	0.174	0.733	0.000	0.000	0.046	0.000	0.000	0.000	0.000	0.000
H_A : Diff. effect size	0.004	0.000	0.000	0.015	0.019	0.031	0.006	0.053	0.002	0.505	0.004	0.000	0.000	0.000	0.000
Baseline mean:															
Boys	0.0443	0.0441	0.0349	0.0301	0.0309	0.0296	0.0274	0.0260	0.0247	0.0267	0.0302	0.0356	0.0385	0.0402	0.0402
Girls	0.0350	0.0349	0.0260	0.0209	0.0204	0.0194	0.0181	0.0167	0.0167	0.0169	0.0160	0.0165	0.0198	0.0268	0.0313
N	2,822,176	3,084,704	3,347,232	3,609,760	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,675,392	3,478,496	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Apr.96	Apr.95	Apr.94	Apr.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93	Jan.93
Latest cohort	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Dec.06	Mar.06	Mar.05	Mar.04	Mar.03	Mar.02

Note: See notes to Table 1. Each regression interacts Sure Start coverage with indicators for whether the cell contains boys or girls (coverage on its own is not included in this model.) ‘Difference p-value’ tests the equality of the coefficients for coverage interacted with boys and with girls. ‘Effect size difference p-value’ tests the equality of the effect size (coefficients weighted by subgroup baseline mean) for coverage interacted with boys and with girls. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

Table A.17: Effect of an increase in Sure Start coverage on probability of hospitalization for mental health, by gender

	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
SS Cov: Boys δ_B	-0.0008***	-0.0016***	-0.0018**	-0.0019
	(0.0003)	(0.0005)	(0.0009)	(0.0013)
SS Cov: Girls δ_G	-0.0006*	-0.0017***	-0.0020**	-0.0000
	(0.0003)	(0.0006)	(0.0010)	(0.0015)
p-values:				
$H_A : \delta_B \neq \delta_G$	0.206	0.802	0.707	0.060
H_A : Diff. effect size	0.143	0.08	0.06	0.026
Baseline mean:				
Boys	0.0013	0.0019	0.0030	0.0035
Girls	0.0013	0.0034	0.0054	0.0062
N	3,215,968	2,953,440	2,690,912	2,428,384
Earliest cohort	Jan-93	Jan-93	Jan-93	Jan-93
Latest cohort	Mar-05	Mar-04	Mar-03	Mar-02

Note: See notes to Table 1. Each regression interacts Sure Start coverage with indicators for whether the cell contains boys or girls (coverage on its own is not included in this model.) ‘Difference p-value’ tests the equality of the coefficients for coverage interacted with boys and with girls. ‘Effect size difference p-value’ tests the equality of the effect size (coefficients weighted by subgroup baseline mean) for coverage interacted with boys and with girls. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

Table A.18: Effect of an increase in Sure Start coverage on probability of hospitalization for any cause, by neighborhood deprivation

	(1) Age 1	(2) Age 2	(3) Age 3	(4) Age 4	(5) Age 5	(6) Age 6	(7) Age 7	(8) Age 8	(9) Age 9	(10) Age 10	(11) Age 11	(12) Age 12	(13) Age 13	(14) Age 14	(15) Age 15
Cov: Poorest 30% δ_P	0.0286***	0.0113	0.0078	-0.0031	-0.0086	-0.0091	-0.0027	-0.0022	-0.0061*	-0.0073**	-0.0102**	-0.0142***	-0.0097*	-0.0091*	-0.0181***
	(0.0093)+++	(0.0079)	(0.0070)	(0.0081)	(0.0071)	(0.0066)	(0.0059)	(0.0043)	(0.0036)	(0.0035)	(0.0041)++	(0.0047)+++	(0.0058)	(0.0049)	(0.0067)++
Cov: Middle 40% δ_M	0.0236***	0.0079	-0.0018	-0.0099*	-0.0142***	-0.0102*	-0.0055	-0.0027	-0.0021	-0.0032	-0.0055	-0.0097**	-0.0058	-0.0102*	-0.0024
	(0.0068)+++	(0.0054)	(0.0046)	(0.0057)+	(0.0051)+++	(0.0053)	(0.0048)	(0.0038)	(0.0029)	(0.0029)	(0.0038)	(0.0039)+	(0.0050)	(0.0052)	(0.0065)
Cov: Richest 30% δ_R	0.0115	0.0013	-0.0013	-0.0060	-0.0120**	-0.0073	-0.0022	0.0001	-0.0007	-0.0004	-0.0011	-0.0021	0.0013	-0.0030	0.0016
	(0.0076)	(0.0057)	(0.0054)	(0.0062)	(0.0057)	(0.0059)	(0.0056)	(0.0045)	(0.0036)	(0.0033)	(0.0040)	(0.0049)	(0.0059)	(0.0070)	(0.0080)
p-values:															
$H_A : \delta_P \neq \delta_M$	0.459	0.561	0.054	0.120	0.125	0.695	0.311	0.855	0.080	0.069	0.063	0.148	0.285	0.813	0.015
$H_A : \delta_P \neq \delta_R$	0.054	0.175	0.135	0.604	0.427	0.630	0.872	0.433	0.046	0.012	0.005	0.002	0.019	0.362	0.011
Effect size diff *p-values*:															
Poor vs. Middle	0.499	0.982	0.048	0.011	0.001	0.090	0.177	0.625	0.194	0.240	0.274	0.577	0.480	0.566	0.037
Poor vs Rich	0.412	0.326	0.208	0.333	0.045	0.601	0.885	0.603	0.175	0.061	0.027	0.016	0.039	0.507	0.033
Baseline mean:															
Poorest 30%	0.3355	0.2708	0.2361	0.2233	0.2138	0.1874	0.1605	0.1472	0.1409	0.1346	0.1304	0.1394	0.1409	0.1499	0.1647
Middle 40%	0.2341	0.1866	0.1645	0.1541	0.1500	0.1333	0.1175	0.1085	0.1051	0.1003	0.1017	0.1098	0.1172	0.1253	0.1338
Richest 30%	0.2027	0.1613	0.1411	0.1332	0.1269	0.1141	0.1027	0.0946	0.0938	0.0908	0.0970	0.1046	0.1123	0.1200	0.1267
N	2822176	3084704	3347232	3609760	3675392	3675392	3675392	3675392	3675392	3675392	3478496	3215968	2953440	2690912	2428384
Earliest cohort	apr.96	apr.95	apr.94	apr.93	jan.93	jan.93	jan.93	jan.93	jan.93	jan.93	jan.93	jan.93	jan.93	jan.93	jan.93
Latest cohort	des.06	des.06	des.06	des.06	des.06	des.06	des.06	des.06	des.06	des.06	mar.06	mar.05	mar.04	mar.03	mar.02

Note: See notes to Table 1. Each regression interacts Sure Start coverage with indicators for whether the LSOA is in the poorest 30%, the richest 30%, or in between (coverage on its own is not included in this model). Indicators for these different groups are time-invariant and so absorbed by the LSOA fixed effects. "p-value: Poor vs. Middle" tests the equality of the coefficients for coverage interacted with the indicators for being in the poorest 30% or the middle. 'p-value: Poor vs. Rich' does the same, testing the equality of the coefficients relating to the poorest and richest 30%. 'Effect size p-value' does the same, but for the effect sizes (coefficients weighted by subgroup baseline mean). *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively; $+$, $++$ and $+++$ indicate significance at the 10%, 5% and 1% level, respectively, after adjusting inference following the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005).

B Data sources for local authority characteristics

This appendix provides further detail on the sources, years of measurement and geographic levels of the local characteristics used in our quantitative analysis of the rollout of Sure Start in section 5.1.1

Table B.1 shows that for most characteristics we have data covering the entire period between 1999 and 2010. A major exception to this is the share of primary school pupils with English as an additional language (where data are not available between 2000 and 2003). In this case, we have imputed the data from these missing years with a constant and included a 'missing' dummy to avoid dropping these observations.

In addition, many of the data series have casewise missingness, where data are unavailable for some area–year combinations (but not more generally for the entire year or for the same area in every year). We use linear interpolation to reduce missingness in these data by imputing the missing data as an average of the non-missing observations in the same area in the year before and after. We apply this procedure in cases where up to five years of data are missing. Within the 323 local authority districts that we consider in the main impact analysis (dropping the City of London, Isles of Scilly and West Somerset, which were all strong outliers in Sure Start coverage), no casewise missing data remain after this procedure.

Table B.1: Covariates used in the rollout analysis

Category	Variable	Source	Years	Geography
Deprivation	Percentile of rank distribution of Index of Local Deprivation	Department of Environment, Trade, and the Regions[1]	1998	LAD
Health Indicators	Under-18 conception rate (conceptions/1,000 women aged 15-17)	Child and Maternal Health Intelligence Network[2]	1998-2018	LAD
	Proportion of births below 2.5kg	ONS Vital Statistics[3]	1991-2018 (interpolated in 2008 and 2009)	LAD
Potential Demand for services	Total period fertility rate	ONS Vital Statistics[4]	1990-2018	LAD
	Density	ONS Population Density[5]	1990-2018	LAD
	% of primary school pupils with English as an Additional Language	National Association for Language Development in the Curriculum (NALDIC)	1999; 2004-2018	County
	Children Looked After per thousand (under 1)	Department for Education	1992-2018	County
	Children Looked After per thousand (1 to 4)	Department for Education	1992-2018	County
Labour Market	Rate of Jobseekers Allowance receipt	Jobseekers Allowance[7]	1992-2018	LAD
Pre-Existing Services	Number of GPs per 1,000 population	Constructed with HSCIC data[10]	1990-2018	LAD
	Number of JobcentrePlus per 1,000 population	Department for Work and Pensions	2001-2018	LAD
	Free entitlement take-up rate among 3 and 4-year-olds	Department for Education Statistical Returns	1997-2018	County

Note: [1] Downloaded 20 Nov. 2015, http://www.legco.gov.hk/yr99-00/english/bc/bc09/papers/1471e01.pdf. [2] Downloaded 02 Nov. 2015, http://atlas.chimat.org.uk/IAS/dataviews/view?viewId=96. [3] Obtained 24 Nov. 2015 from the ONS Vital Statistics Outputs Branch, with help from Laura Todd. [4] Obtained 24 Nov. 2015 from the ONS Vital Statistics Outputs Branch, with help from Laura Todd. [5] Downloaded 18 January 2016 from ONS. [6] Downloaded 02 Dec. 2015 from NOMIS. [7] Downloaded 16 Dec. 2015 from NOMIS. [8] Deflated to constant 2015 pounds using the Consumer Price Index, downloaded from ONS Consumer Price Indices – Tables, table 1.1, series CPI All Items Index (estimated pre-97, 2005=100) on 27 January 2016. http://www.ons.gov.uk/ons/datasets-and-tables/data-selector.html?dataset=mm23. [9] Downloaded 15 December 2015 from NOMIS. [10] HSCIC, 'GPs, GP Practices, Nurses, and Pharmacies', downloaded 26 November 2015.

C Evidence on migration between local areas

In our main analysis, we assess how early exposure to Sure Start in a child's local authority of residence affects the probability of hospitalization between ages 1 and 15. We define children's local authority based on their residence at the time of hospitalization, since residence at the time of birth is not reliably measured for cohorts born before 2003.

There are two potential difficulties in using a child's residence-at-admission as the basis for defining their exposure to Sure Start. First, mobility across local authorities could introduce measurement error if we assign children's treatment based on the wrong local authority. Second, to the extent that mobility is selective (for example, with more motivated families electing to move into areas with greater access to Sure Start), our strategy will yield biased estimates of Sure Start's effectiveness.

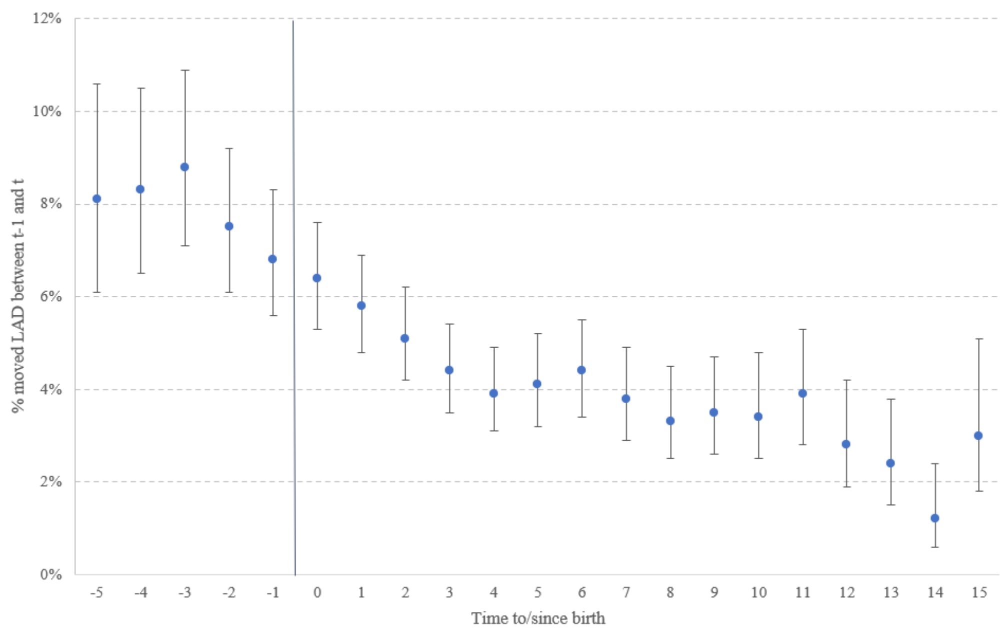
In this appendix, we use data restricted-access data with geographic identifiers from the British Household Panel Survey (BHPS) to assess both the overall extent of inter-LA mobility during childhood and the extent to which it is correlated with Sure Start provision. The BHPS data are ideal for this analysis: they follow a representative panel annually for 18 years, meaning that we can observe families' mobility before the birth of their child as well as afterwards. Our sample consists of primary caregivers who had a child while in the BHPS sample. We then follow these primary caregivers (henceforth parents) up to five years before the child's birth, and up to 15 years after birth.

Figure C.1 shows that overall inter-LA mobility is relatively low and declining as children age: around 7-8% of families move LA each year in the five years before their child is born, but this declines to 4% of families moving by the time a child is aged 3. This means that measurement error related to mobility between LAs is relatively small, particularly after children turn 5 and age out of Sure Start eligibility entirely.

We also find that what inter-LA mobility there is does not systematically relate to Sure Start availability. In Table C.1, we show that children living at time t in LAs with greater access to Sure Start are no more likely to have moved between t-1 and t than those living in lower-coverage areas.

This provides reassurance that families are not systematically relocating into high-Sure Start local authorities.

Figure C.1: Share of families who moved LA in the past year, by age of child



Note: Mobility is indexed based on the wave in which the family's first birth was observed. Source: Authors' calculations using data from the British Household Panel Survey, 1991-2009.

Table C.1: Association between Sure Start coverage and inter-LA mobility in the previous year

	(1) Age 0	(2) Age 1	(3) Age 2	(4) Age 3	(5) Age 4	(6) Age 5
SS Coverage (at time t)	0.037	0.000	0.020	0.004	0.000	0.003
	(0.041)	(0.026)	(0.021)	(0.019)	(0.018)	(0.017)
Observations	1,017	1,106	1,193	1,134	1,060	1,004
R-squared	0.002	0.000	0.001	0.000	0.000	0.000
Outcome mean	0.072	0.062	0.055	0.046	0.043	0.043

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. The outcome is an indicator for whether the family had moved LA since the previous wave. Sure Start coverage is measured contemporaneously in the LA of residence at time t. Source: Authors' calculations using data from the British Household Panel Survey, 1991-2009.

D Evidence on the relationship between children's behavioral and emotional difficulties and hospitalization

In section 6, we show that Sure Start substantially reduces hospitalizations for accidents and injuries between ages 1 and 11. The impacts at younger ages may be driven by information and support in reducing risk in the home environment; indeed, we find that poisonings fall up to age 3. However, these direct informational effects are not a plausible mechanism for the longer-term falls in external hospitalizations that we observe. In this appendix, we present evidence from the Millennium Cohort Study (MCS) on the correlates of parent-reported injuries, highlighting that child behavior - and particularly externalizing behavior - is strongly associated with injuries through middle childhood and early adolescence.

We use data from three waves of the MCS: wave 3 (age 5), wave 4 (age 7) and wave 5 (age 11).[1] At each wave, parents report injuries sustained by their child since the last wave (so 'age 5' results consider hospitalizations between ages 4 and 5, 'age 7' results for ages 6 and 7, and 'age 11' results for ages 8–11). We use as an outcome whether the parent reports any injury since the previous wave.[2]

The richness of the MCS data allows us to consider the link between child behavior and injuries while controlling for a wide range of other potential correlates. These include child demographics (sex, ethnicity); maternal demographics (age at child's birth, education); economic circumstances (maternal employment, household net earnings); and the home learning environment.[3] In our specification with covariates, we also control for fixed effects for the child's region of residence at the time of interview[4] and the season of the interview. All regressions control for fixed effects of

[1]The MCS data are taken from a single cohort born in 2000–01 and so will not necessarily be representative of all the cohorts in our impact analysis.

[2]The associations documented here are similar when we consider the number of injuries sustained since the previous wave. We focus on the 'any injury' indicator since it is more analogous to the outcomes in our main results.

[3]We construct a standardized measure of the home learning environment using factor analysis on a series of parental time inputs (for example, how often the parents read to the child, visit the library, or play games with the child).

[4]There are 12 regions in total: 9 in England, Wales, Scotland and Northern Ireland. Since regional data is not available in the public-access version of MCS at wave 4 (age 7), we use the child's wave 3 (age 5) region of residence instead.

the child's age in months at the time of interview.

We have two measures of child behavior, both reported by the mother through the widely used Strengths and Difficulties Questionnaire (SDQ). The 'externalizing behavior' score comes from the SDQ subscales on hyperactivity and conduct disorders. The 'internalizing behavior' score comes from the SDQ subscales for emotional problems and peer problems. Both indices are scored out of 20, with a higher score indicating more problems in that domain of behavior.

Table D.1 presents the associations between behavior and whether the child has sustained an injury since the previous wave. Columns 1, 3 and 5 show a significant relationship between externalizing behavior problems and injuries at ages 5, 7 and 11. An additional point on the externalizing scale (out of 20) is associated with roughly a roughly 1-percentage point increase in the probability that a child has sustained an injury since the previous wave. By contrast, there is little association between internalizing behavior and injuries at any age.

Columns 2, 4 and 6 show that these associations are robust to the inclusion of a wide range of additional controls. While these results should not be interpreted as causal, they do provide suggestive evidence of a relationship between externalizing behavior and injuries that cannot be explained by the child's demographics or family circumstances. This supports the hypothesis that a Sure Start-induced change in child behavior is a plausible mechanism for the reduction in injury-related hospitalizations through middle childhood and early adolescence. This is also in line with findings from the ECCE project, which identified a reduction in externalizing behavior over time as one of the main benefits of using Sure Start services (Sammons, Goff and Smith, 2015).

Table D.1: Association between child behavioral problems and any parent-reported injury

	(1)	(2)	(3)	(4)	(5)	(6)
	Age 5		Age 7		Age 11	
Externalizing behavior	0.009***	0.006***	0.009***	0.007***	0.011***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Internalizing behavior	0.001	0.002	-0.001	0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female		-0.043***		-0.033***		-0.032***
		(0.011)		(0.010)		(0.012)
Ethnicity: Mixed		-0.052*		-0.038		-0.055*
		(0.029)		(0.028)		(0.031)
Ethnicity: Indian		-0.043		-0.056**		-0.180***
		(0.028)		(0.027)		(0.031)
Ethnicity: Pakistani or Bangladeshi		-0.088***		-0.095***		-0.179***
		(0.021)		(0.020)		(0.023)
Ethnicity: Black		-0.075**		-0.043		-0.147***
		(0.030)		(0.029)		(0.032)
Ethnicity: Other		-0.068*		-0.054		-0.203***
		(0.040)		(0.038)		(0.043)
Mother's age at birth		-0.001		0.000		-0.000
		(0.001)		(0.001)		(0.001)
Mother cohabiting		-0.035**		-0.012		-0.024
		(0.015)		(0.014)		(0.015)
Mother's education: A level		0.010		-0.002		-0.004
		(0.017)		(0.016)		(0.018)
Mother's education: GCSE or below		0.017		0.016		-0.006
		(0.013)		(0.012)		(0.014)
Mother's education: Missing		0.034		-0.003		-0.043
		(0.025)		(0.025)		(0.030)
Mother's work status: Part-time		0.005		-0.009		-0.018
		(0.013)		(0.013)		(0.015)
Mother's work status: Full-time		0.001		0.021		-0.015
		(0.014)		(0.013)		(0.016)
Mother's work status: Unknown		-0.109*		0.017		-0.055
		(0.061)		(0.052)		(0.044)
Household net earnings		0.000		-0.000		-0.000
		(0.000)		(0.000)		(0.000)
Home learning environment		0.011*		0.012**		0.018***
		(0.006)		(0.005)		(0.006)
Constant	0.219***	0.320***	0.190***	0.246***	0.315***	0.469***
	(0.009)	(0.044)	(0.008)	(0.042)	(0.009)	(0.047)
Observations	6,971	6,960	6,974	6,963	6,947	6,924
R-squared	0.007	0.019	0.009	0.020	0.010	0.036
Mean	.26	.26	.228	.228	.358	.358
Age in months FE?	Yes	Yes	Yes	Yes	Yes	Yes
Region FE?		Yes		Yes		Yes
Interview quarter FE?		Yes		Yes		Yes

Note: The outcome is an indicator for whether the parent reports that the child has sustained an injury since the last wave (i.e. from ages 3-5 for age 5 results; from ages 6-7 for age 7 results; and from ages 8-11 for age 11 results). *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

E Wider outcomes and mechanisms

The analysis in this paper has focused on the impact that access to Sure Start has on children and adolescents' hospitalizations. As a measure of serious - and costly - health problems, hospitalizations are an important outcome to consider. However, changes in hospitalizations could reflect changing patterns of healthcare usage as well as changes in underlying health. In this section, we therefore consider the impact that Sure Start had on the self-reported physical and mental health of young people and on infant mortality.

E.1 Self-reported physical and mental health

The UK Household Longitudinal Study (UKHLS) is a long-standing survey panel in the UK. This survey covers a representative sample of around 40,000 households, with annual interviews starting in 2010. Adolescents in the household (aged 11 to 15) self-complete a dedicated survey covering topics including mental health (as assessed by the Strengths and Difficulties Questionnaire) and self-reported health.

To identify the impact of Sure Start on young people's health outcomes, we adopt a similar approach to the one described in subsection 4.2. Specifically, we regress outcomes in adolescence on a young person's exposure to Sure Start, as measured by the average number of centers per thousand children in their local area over the first five years of life.

For identification, we take advantage of the survey's coverage of all people in the household to implement a family fixed effects strategy, exploiting variation in access to Sure Start across siblings. By exploiting within-household variation, this strategy removes the influence of observed and unobserved household-level characteristics, including characteristics of the local area, the home environment, or parental preferences for services like Sure Start. Specifically, we estimate:

$$D_{imda}^{y} = \delta^{y} SS_{dq} + \pi_{m}^{y} + \phi_{q}^{y} + \psi_{t}^{y} + \gamma_{a}^{y} + \beta^{y} X_{it} + \epsilon_{imda}^{y} \tag{5}$$

where D_{imda}^{y} is the outcome at age a of child i of mother m, living in local authority d. SS_{dq}

is our measure of Sure Start coverage, defined as the average over the first 20 quarters after birth in quarter q of the number of centers per thousand children aged 0-4 in the young person's local authority d. π^y_m is a set of mother fixed effects, which allow us to identify the impact of SS_{dq} as it varies between siblings. We further control for a range of additional fixed effects for the year-quarter of birth (ϕ^y_q), for the quarter of interview (ψ^y_t), and for the young person's age in years at interview (γ^y_a). Finally, we control for several additional characteristics in X_{it}. These include the young person's gender, ethnicity, birth order and whether the household contains any other child aged 0-5, 6-10, 11-15, 16-19 or 20+. Of these, the control for birth order is particularly important, since the expansion of Sure Start over our study period means that later-born children will always experience weakly more treatment than their older siblings.

We impose similar sample restrictions to our main hospitalization estimates; we include young people born between 1993 and 2006 and residing in England at the time of interview (excluding the three 'outlier' local authorities with exceptionally high Sure Start coverage due to small populations). Finally, we limit the sample to families with multiple children to allow us to implement our family fixed effects strategy.

Results In Table E.2, we present the results of an analysis of five outcomes. The first two columns show the impact of greater Sure Start coverage on young people's overall mental health, as measured by the Strengths and Difficulties Questionnaire (SDQ). The SDQ measures socio-emotional and behavioral problems, so higher scores correspond to greater problems and worse mental health. We see from the table that young people who had greater access to Sure Start in their early childhood score around 3.5 points lower on the SDQ than their less-exposed siblings. This is a significant improvement: at baseline, the mean SDQ score was just over 11 points, so this represents a more than 30% improvement. The effect remains in Column 2, when we add in linear time trends interacted with baseline measures of the determinants of Sure Start's rollout as a robustness check. These results validate our findings in section 5, suggesting that the reduction in mental health-related hospitalizations among adolescents was indeed related to improvements in mental health, rather than changes in seeking mental health care.

We can further explore this overall improvement in mental health by distinguishing between ‘internalizing’ behaviors (such as depression or anxiety) and ‘externalizing’ behaviors (including aggression and hyperactivity). While the point estimates suggest improvements in both dimensions, it is clear from Table E.2 that Sure Start predominantly reduces internalizing behavior. While the lack of data on the services offered by each center means we cannot pin down which types of support drive these benefits, this improvement is consistent with improvements in parenting and the home environment supporting children’s social and emotional development. Further analysis in columns 9 and 10 gives suggestive (though not significant) evidence that Sure Start exposure improves young people’s relationship with their parents, making them more likely to report feeling supported by their family most or all of the time.

The final set of columns in Table E.2 report the impact of greater access to Sure Start on young people’s self-reported health, specifically a measure of whether they report having very good or excellent health. Unfortunately, objective measures of physical health like health conditions are not available for this age group; this means we must rely on this self-reported measure, which in principle captures both physical and mental health. We see that Sure Start significantly and substantially increases the proportion of young people reporting very good or excellent health: a one-unit increase in coverage raises the share of young people with very good or excellent health by 18 percentage points, around a third of the baseline mean.

Table E.2: Effect of an increase in Sure Start coverage on young people's mental and self-reported health

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	SDQ: Total		SDQ: Internalising		SDQ: Externalising		V. good or excellent health		Supportive family	
Sure Start coverage	-3.474***	-3.080**	-2.502***	-2.720***	-0.960	-0.841	0.182**	0.183**	0.127	0.120
	(1.259)	(1.294)	(0.722)	(0.706)	(0.858)	(0.851)	(0.085)	(0.085)	(0.086)	(0.085)
Constant	10.578***	10.524***	4.787***	4.922***	5.766***	5.812***	-0.030	-0.033	1.022***	1.032***
	(0.906)	(0.855)	(1.517)	(1.264)	(1.852)	(1.910)	(0.090)	(0.089)	(0.060)	(0.059)
N	8,192	8,192	8,197	8,197	8,195	8,195	8,047	8,047	8,221	8,221
N children	4,667	4,667	4,670	4,670	4,667	4,667	3,974	3,974	4,691	4,691
N families	2,066	2,066	2,067	2,067	2,066	2,066	1,761	1,761	2,076	2,076
Family FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
LA trends * 1998 rollout		Y		Y		Y		Y		Y
Baseline mean	11.225	11.225	4.544	4.544	6.669	6.669	0.522	0.522	0.761	0.761
Baseline SD	5.443	5.443	3.031	3.031	3.659	3.659	0.501	0.501	0.427	0.427

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. The table reports estimates of the coefficient associated with Sure Start exposure in an OLS regression of the dependent variable indicated at the top of each column, pooling outcomes measured between ages 11 and 15. Each specification includes family fixed effects as well as indicators for the year-quarter of birth, the quarter of interview, the age at interview, ethnicity, birth order and whether the household contains any other children aged 0-5, 6-10, 11-15, 16-19 or 20+. The second column in each pair additionally controls for linear trends interacted with baseline (1998) characteristics of the local authority, including its Index of Local Deprivation rank, the share of births with low birth weight and the teen conception rate. Standard errors clustered at the Local Authority level.

E.2 Infant and child mortality

We have also estimated the impacts of Sure Start on infant and child mortality rate. However, there are some data limitations. In particular, LSOA level mortality data is only available at yearly level and from 2002, and therefore unavailable during most of the years of SSLP - 1999-2002; this is because LSOAs were created in 2001. Nevertheless, we re-estimated model (1) at yearly level using data on infant and child mortality for the period of 2002-2017 and data for each cell LSOA-sex-year. We then control for gender, LSOA and year fixed effects and contemporaneous LSOA population in the cell. We find no association between the number of centers per thousand 0-4 children available in the LSOA at the year of birth and infant mortality. This suggests that the drop in infant mortality rate in England from 5.2 per 1,000 live births in 2002 to 3.7 in 2019 is not due to the Sure Start expansion.[5] There is also no association between child mortality (1 to 4 years old) and number of centers per thousand 0-4 children available in the LSOA in a given year (see

[5] https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/childhoodinfantandperinatalmortalityinenglandandwales/2019.

Table E.3).

These results do not seem to be driven by the sampling period, since we find also no association between infant and child mortality rates and the expansion of Sure Start using LA level mortality data that is available from 1998 (the earliest year for which such data is collected by the Office for National Statistics at LA level). To be precise, we use LA-year data and estimate a version of model (1), that includes the number of centers per thousand 0-4 children available in the LA, LA and year fixed effects, weighting by contemporaneous LA population in the cell.

Table E.3: Impact on Mortality Rate

	(1) IMR	(2) Child MR
Coverage SS 0-4	199.781	-12.534
	(217.475)	(16.148)
Mean	6.075	0.233
N	1044908	1049639

Note: Data source: Mortality data 2002-2017. Mortality by age group in an LSOA-sex-year cell (before age 1 in column 1 and between ages 1 and 4 in column 2). The controls included in the model but excluded from the tables are sex, LSOA and year of birth fixed effects and contemporaneous population in the cell. The mean is taken in 2002. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

F Estimation of Sure Start effects on maternal employment

In addition to their focus on children's health and development, Sure Start centers also brought together existing services to support parental employment. Children's centers were required to develop links with JobcenterPlus, an existing network of government-run agencies to support the unemployed in finding work. Children's centers were also required to signpost parents towards existing childcare programs, most notably the entitlement to a part-time free childcare place for 3- and 4-year-olds.[6] Many Sure Start centers also offered information about further education and basic skills courses.

There is a large literature establishing that childcare subsidy programs can affect parental employment in some contexts, but typically only for mothers whose youngest child is eligible for the program (e.g. Gelbach (2002); Cascio (2009); Brewer et al. (2020)). While these parental employment outcomes are important in their own right, an increase in parental employment may also impact on children's development through higher family income and/or less parental time with children. To investigate the likely importance of this channel, we use the UK's Labour Force Survey (LFS) to analyze how access to Sure Start affected maternal employment.

Data and outcomes The LFS is collected in a staggered five-quarter rolling panel, with households entering the survey at different points in the year and then remaining in the sample for five consecutive quarters. We use a secure access version of the LFS that contains information both on the household's local authority of residence and on the precise birth date of all household members. To mirror our hospitalization analysis, we focus on mothers whose children were all born between 1993 and 2006. To avoid mothers of newborn children (who most often take several months of maternity leave), we further restrict the sample to mothers who did not give birth during the period that they were in the LFS sample.

As our primary outcome, we focus on an indicator for whether a mother is in paid work at the

[6] The free entitlement was first introduced in 1997, offering a free childcare place to 4-year-olds for 12.5 hours per week, 33 weeks of the year. The program was extended to cover 3-year-olds in April 2004, and the generosity of the system was increased in a series of reforms: by September 2010 it covered a 15-hour place for 38 weeks of the year.

time she is surveyed by the LFS. As secondary outcomes, we consider whether mothers work part-time (fewer than 30 hours a week) or full-time, and whether they are in full-time education. Since mothers are observed up to five times in the LFS, each mother can be included multiple times in our model.

Sure Start treatment Since existing evidence suggests that the strongest effects should be found among mothers whose youngest child is eligible for support, we focus on the treatment a mother experiences in respect of her youngest child. Specifically, we use the same measure of Sure Start coverage as in our hospitalization analysis (centers per thousand children aged 0-4 in the local authority, averaged over the child's first five years of life[7]). We assign this measure of Sure Start coverage to mothers based on the year and month of birth of their youngest child and their local authority of residence when they are first observed in the LFS.

Specification To evaluate the impact of access to Sure Start on maternal employment, we estimate Equation 6 by OLS:

$$y^a_{iwat} = \alpha + \delta^a SS_{dq} + \pi_w + \lambda_t + \gamma^{a_m}_{it} + \phi^{a,k} g^k_{it} * Ki\beta^a X_i + \epsilon^a_{wdmt}, a = 0, ..., 15 \tag{6}$$

where y^a_{iwat} is the outcome variable, an indicator for whether a mother i living in ward w is in work in year-quarter t when her youngest child is a years old. We estimate the model separately for each age of the youngest child from 0 to 15. SS_{dq} is the average Sure Start coverage of the mother's youngest child, based on the year-quarter of birth q and where the family resides when they enter the LFS (the local authority d that contains ward w). We include year-quarter fixed effects t to control for contemporaneous labor market conditions. We control flexibly for the ages of children in the household: $\gamma^{a_m}_{it}$ is a set of fixed effects for the youngest child's age in months at the time mother i is observed in year-quarter t. We also control for the presence and ages of up to four older children k through a continuous measure of the older child's age in years g^k_{it}, interacted with an

[7]Where a child is less than five years old, we average coverage only over the quarters in which they have actually been alive.

indicator K_i for whether there is such a child in the household.

Unlike our main hospitalization regressions, Equation 6 is estimated at the individual level. This means we are able to control for individual characteristics X_i. We include characteristics pre-determined at the time of potential Sure Start exposure, namely mothers' ethnicity and age; in alternate specifications we also include education and partnership status. However, because Equation 6 is estimated on individual-level survey data, we cannot include LSOA-level fixed effects (since there are not sufficient observations in each LSOA). We instead control for around 9,000 ward fixed effects (π_w).

F.1 Results: Maternal employment

We first consider the impact that Sure Start had on the probability that a mother is working. These results are presented in Figure F.1, which reports the estimates from 15 separate regressions, based on the age of the mother's youngest child. To account for the different baseline probabilities of employment at different ages, Figure F.1 then rescales each of these coefficients by the baseline employment rate of women whose youngest child was that age in 1996.

Figure F.1 shows no clear pattern in Sure Start's impacts on maternal employment. While there are statistically significant positive impacts at ages 1, 6 and 15 (and a significant negative effect at age 7), there is no clear overall pattern of results across ages. We present the full set of results in Column 2 of Tables F.4 to F.6.[8]

Column 1 presents the raw correlation between Sure Start coverage and maternal employment. Unsurprisingly, mothers with greater access to Sure Start - whose children are on average older - tend to have higher rates of employment.[9] In Column 2 we control for ward fixed effects and for the set of basic controls shown in Equation 6. In the next three columns we present additional robustness checks. In Column 3 we allow for a local authority-specific linear time trend, estimated

[8]We conduct similar analysis for subgroups of mothers: single mothers, partnered mothers, and by maternal education (those with less than high school vs. mothers with high school or more). We find no consistent patterns of impacts among any of these subgroups. Results available on request.

[9]This is because Sure Start treatment is generally weakly increasing over a child's first five years, as new centers open in the child's local authority. Therefore, as children who are still age-eligible for Sure Start get older, their average level of access to Sure Start tends to increase.

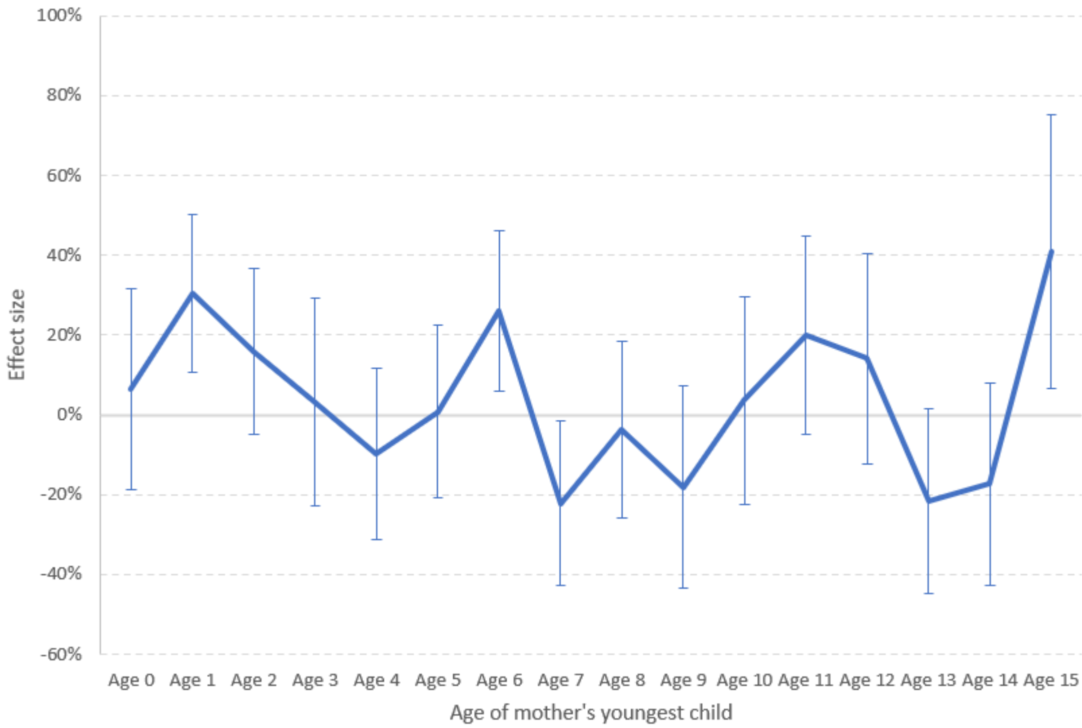
based on pre-treatment data and extrapolated to the post-Sure Start period. The inclusion of these estimated trends has very little impact on our results. In Column 4 we additionally control for characteristics of the mother that were potentially influenced by Sure Start exposure (education and partnership status); characteristics of the local labor market at the time of data collection (male and female median weekly full-time earnings and the local unemployment benefit claiming rate); and a range of local characteristics that may have helped to determine Sure Start's rollout, measured at the birth of the youngest child.[10] In general, the inclusion of this extended set of controls does not change the overall conclusion of mixed impacts of Sure Start on maternal employment, with mostly non-significant effects.

As a final robustness check, in Column 5 we estimate our main equation (Column 2) on the subgroup of mothers with only one child. This sample restriction allows us to examine maternal employment in the simplest case, without the possibility of unmeasured spillovers from older children's treatment. Our results become substantially less precise, but we find similar patterns in terms of the direction and statistical significance of effects, except at the oldest ages.

We also present the results of a specification check in Column 6. Here, we exploit the panel aspect of the LFS to control for mother fixed effects. This allows us to look within mothers at whether higher Sure Start coverage increases the probability that a mother is working. Because Sure Start coverage only varies during a child's first five years of life (as the average coverage is updated to include additional months of treatment), this specification is only possible where the youngest child is aged 4 or below (Table F.4). This specification substantially decreases the precision of our estimates, but again we find statistically significant employment impacts only at age one. These effects are very large - implying that a mother gaining an additional center per thousand children was nearly 30 percentage points more likely to be working - but they once again come in a context of insignificant and inconsistent results at other ages.

[10]This is the same set of characteristics used in the robustness checks for our hospitalization analysis.

Figure F.1: Effect of Sure Start coverage on probability of maternal employment, rescaled by baseline probability



Note: The table shows coefficients from separate regressions for each outcome age. Coefficients are rescaled by the employment rate of mothers whose youngest child was born in 1996. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the UK Labour Force Survey and the Department for Education's data on the rollout of Sure Start.

Table F.4: Effect of an increase in Sure Start coverage on probability of maternal employment: Youngest child aged 0-4

	(1)	(2)	(3)	(4)	(5)	(6)
Age 0	-0.014	0.032	0.022	0.005	0.033	-0.139
	(0.022)	(0.077)	(0.089)	(0.085)	(0.102)	(0.130)
N	28,190	28,190	28,190	28,190	16,087	28,190
Baseline mean	0.5036	0.5036	0.5036	0.5036	0.5349	0.5036
Age 1	-0.044**	0.165**	0.150**	0.147**	0.280***	0.285**
	(0.020)	(0.065)	(0.073)	(0.074)	(0.084)	(0.128)
N	45,595	45,595	45,595	45,595	25,147	45,595
Baseline mean	0.5429	0.5429	0.5429	0.5429	0.5883	0.5429
Age 2	-0.063***	0.087	0.05	0.091	0.007	0.105
	(0.021)	(0.069)	(0.068)	(0.068)	(0.119)	(0.142)
N	37,605	37,605	37,605	37,605	19,065	37,605
Baseline mean	0.5449	0.5449	0.5449	0.5449	0.5825	0.5449
Age 3	-0.063***	0.018	0.02	0.028	-0.028	-0.055
	(0.020)	(0.091)	(0.090)	(0.095)	(0.149)	(0.175)
N	31,162	31,162	31,162	31,162	14,282	31,162
Baseline mean	0.5774	0.5774	0.5774	0.5774	0.6178	0.5774
Age 4	-0.070***	-0.063	-0.063	-0.046	-0.206	0.166
	(0.018)	(0.083)	(0.083)	(0.081)	(0.158)	(0.186)
N	27,028	27,028	27,028	27,028	11,473	27,028
Baseline mean	0.6411	0.6411	0.6411	0.6411	0.6732	0.6411
Fixed effects		Ward	Ward	Ward	Ward	Mother
Trends?			Estimated	Estimated	Estimated	
Basic Controls?		Y	Y	Y	Y	Y
Extended Controls?				Y		
Sample restrictions?					Only children	

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. Standard errors are clustered at the LA level.

Table F.5: Effect of an increase in Sure Start coverage on probability of maternal employment: Youngest child aged 5-10

	(1)	(2)	(3)	(4)	(5)
Age 5	-0.092***	0.005	0.009	-0.047	-0.159
	(0.016)	(0.092)	(0.090)	(0.089)	(0.203)
N	24,247	24,247	24,247	24,247	9742
Baseline mean	0.7013	0.7013	0.7013	0.7013	0.69
Age 6	-0.063***	0.183**	0.179**	0.221**	-0.006
	(0.017)	(0.086)	(0.089)	(0.089)	(0.204)
N	22,292	22,292	22,292	22,292	8727
Baseline mean	0.7039	0.7039	0.7039	0.7039	0.7028
Age 7	-0.034**	-0.165*	-0.161*	-0.122	0.199
	(0.017)	(0.093)	(0.093)	(0.091)	(0.228)
N	21,148	21,148	21,148	21,148	8247
Baseline mean	0.7453	0.7453	0.7453	0.7453	0.7045
Age 8	-0.014	-0.028	-0.024	-0.134	0.247
	(0.017)	(0.101)	(0.101)	(0.111)	(0.324)
N	20,610	20,610	20,610	20,610	7956
Baseline mean	0.7487	0.7487	0.7487	0.7487	0.7265
Age 9	-0.022	-0.137	-0.118	-0.185	-0.377
	(0.018)	(0.116)	(0.119)	(0.136)	(0.259)
N	19,834	19,834	19,834	19,834	7538
Baseline mean	0.7571	0.7571	0.7571	0.7571	0.738
Age 10	0.004	0.028	0.000	0.059	0.156
	(0.017)	(0.123)	(0.123)	(0.127)	(0.271)
N	19,116	19,116	19,116	19,116	7228
Baseline mean	0.7795	0.7795	0.7795	0.7795	0.7906
Fixed effects		Ward	Ward	Ward	Ward
Trends?			Estimated	Estimated	Estimated
Basic Controls?		Y	Y	Y	Y
Extended Controls?				Y	
Sample restrictions?					Only children

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. Standard errors are clustered at the LA level.

Table F.6: Effect of an increase in Sure Start coverage on probability of maternal employment: Youngest child aged 11-15

	(1)	(2)	(3)	(4)	(5)
Age 11	0.012	0.157	0.152	0.178	0.113
	(0.018)	(0.119)	(0.120)	(0.139)	(0.297)
N	18,784	18,784	18,784	18,784	7357
Baseline mean	0.7886	0.7886	0.7886	0.7886	0.7794
Age 12	-0.007	0.113	0.112	0.008	-0.01
	(0.018)	(0.128)	(0.126)	(0.122)	(0.290)
N	18,809	18,809	18,809	18,809	7651
Baseline mean	0.8007	0.8007	0.8007	0.8007	0.8316
Age 13	0.014	-0.174	-0.169	-0.028	-0.582*
	(0.017)	(0.113)	(0.111)	(0.112)	(0.311)
N	17,854	17,854	17,854	17,854	7650
Baseline mean	0.806	0.806	0.806	0.806	0.7816
Age 14	-0.026	-0.14	-0.159	-0.072	-0.657**
	(0.023)	(0.124)	(0.121)	(0.124)	(0.302)
N	16,385	16,385	16,385	16,385	7829
Baseline mean	0.8053	0.8053	0.8053	0.8053	0.769
Age 15	-0.014	0.324*	0.348**	0.459***	0.213
	(0.028)	(0.165)	(0.164)	(0.177)	(0.265)
N	14,835	14,835	14,835	14,835	8064
Baseline mean	0.7932	0.7932	0.7932	0.7932	0.7613
Fixed effects		Ward	Ward	Ward	Ward
Trends?			Estimated	Estimated	Estimated
Basic Controls?		Y	Y	Y	Y
Extended Controls?				Y	
Sample restrictions?					Only children

Note: *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively. Standard errors are clustered at the LA level.

G Cost-Benefit Analysis of Sure Start

This section reports the details of the cost–benefit calculation we report in section 8 of the paper. We do so by combining official data on government expenditures on Sure Start to compute the cost of Sure Start, with the estimates obtained in the previous sections and results from the best published literature to compute the benefits. We compute the averted costs in terms of hospitalizations attributable to providing access to Sure Start to 1,000 more children (i.e. from opening one more center at the peak coverage level).

We are not the first to try to quantify the monetary benefits of Sure Start. Meadows (2011) calculated that SSLPs cost around £1,300 per eligible child per year at 2009–10 prices (or £4,860 per eligible child over the period from birth up to age 4); and that by the time children had reached the age of 5, SSLPs had already delivered economic benefits between £279 and £557 per eligible child (coming from reduction in work-less households), which is 6–12% of the total cost of the program. The authors concluded that this is a large impact, given the early stage at which it is measured, but that there was insufficient information to reliably predict longer-term economic impacts.

Gaheer and Paull (2016) collected very detailed cost data on different types of services delivered in 24 of the SSCCs that participated in the ECCE: baby health, child play, parent support, specialist child support, specialist family/parent support, childcare, finance and work support, and training and education. The average cost per user per hour (the value of resources used to deliver one hour of a service to a child) ranged from £6 for childcare to £55 for finance and work support, while the mean cost per family using the service (which accounts for the hours of usage) ranged from £958 for parent support to £8,454 for childcare. The authors then combined estimates on the associations between the use of different types of SSCC services and improved family outcomes with existing evidence from the literature on long-term effects. They found that some SSCC services provide positive value for money, i.e. the monetary valuation of improved outcomes exceeds the cost of delivery.

Costs We opted to compute in an alternative way the cost of Sure Start. Our choice is informed by different factors. First, we have not collected detailed costs data as was done in the NESS and ECCE evaluations. Second, given that we evaluate the effects of Sure Start using the whole period it was in place, it would be difficult to compute a measure of costs valid for both SSLPs and SSCCs. Third, our measure of costs needs to be consistent with the methods we use in the estimation of the impacts, which studies the effects of access to, rather than usage of, Sure Start. For these reasons, we compute the cost of Sure Start per eligible child, dividing the overall government expenditures on Sure Start by the number of eligible children, i.e. the number of children aged 0–4 in the local authorities in which Sure Start was in place in that particular year. This is consistent with the aim of the government (especially at program maturity) to provide Sure Start to every age eligible child, and the fact that Sure Start was area-based, rather than means-tested. The cost per child computed in this way amounts to £415.9 per eligible child, on average.[11]

Benefits Weighed against Sure Start's cost to taxpayers, we consider the financial benefits of the hospitalizations that Sure Start averted. In doing this calculation, we only consider impacts that are statistically significant at the 10% or less after accounting for multiple hypothesis testing, for the following conditions: injuries and poisoning (a subset of external), respiratory, parasitic/intestinal, and mental health. We consider three types of costs:

- Averted direct healthcare costs. We use specific NHS resource use costs for each of these conditions, taking the average cost among the different categories for non-elective long and short stay.
- Averted indirect costs, over the same ages as the healthcare costs, such as costs to the family and to society (e.g. lost income and value of work time lost).
- Averted long-term costs, for those cases that would incur sustained costs over the life cycle (such as those deriving from traumatic brain injury or attributable to child maltreatment, or

[11] Although information on Sure Start usage is scare, we can also use the Action for Children (2019) estimate of 2.2 million yearly users in 2013 to compute the cost per child using the services, which amounts to around £480.

for mental health conditions).

The main results of our cost–benefit calculation are reported in Table G.1. All costs are in 2018–19 prices, and discounted using a 3.5% discount rate as recommended by the National Institute for Health and Care Excellence (NICE). The total financial benefit from averted costs, obtained by adding together the direct healthcare costs, indirect costs throughout childhood and long-term costs, amounts to around £330 million. Of this, around £3.9 million is attributed to direct cost savings to the NHS from fewer hospitalizations at ages 1–15. As expected, the bulk of the total averted cost is attributable to the lifetime costs of traumatic brain injury and mental health conditions. Set against this is the estimated cost of providing an additional Sure Start center per thousand children to a representative cohort, which we calculate at £1,055 million. On this basis, then, we find that the financial benefits from reducing hospitalizations offset approximately 31% of the cost of Sure Start provision (with direct savings from the reduction in hospitalizations at ages 1–15 amounting to 0.37% of spending on Sure Start).

Table G.1: Estimated costs and benefits of Sure Start for one cohort of children (2018–19 prices)

Total program expenditures	£1,055 million
Total costs from averted hospitalizations	£330 million
Of which:	
Direct healthcare costs (1.2%)	£3.9 million
Indirect costs (1.3%)	£4.3 million
Long-term costs (97%)	£322 million

The total averted costs of Sure Start were calculated using the estimated effect of Sure Start on hospital admissions for poisoning, head injuries, fractures, respiratory illnesses, infections and parasitic conditions and mental health. The direct healthcare costs are calculated using the National Schedule of NHS Costs (2018/2019). To compute the indirect costs, we use Cooper et al. (2016)'s estimated mean short term family costs resulting from injury and poisoning hospitalizations; Stevens et al. (2003)'s family borne cost of respiratory admissions and Telford et al. (2012)'s estimated mental health educational cost. The sources used to calculate the lifetime costs of averted hospitalizations are the following:

- We compute the share of head injuries and fracture hospital admissions being due to child maltreatment using González-Izquierdo et al. (2010); and calculate their lifetime costs based on Conti et al. (2017).
- The proportion of traumatic brain injury admissions is calculated using Trefan et al. (2016). The medical and lifetime costs of a pediatric traumatic brain injury are based on Kendrick et al. (2017) and Child Accident Prevention Trust (2013), respectively.
- We use Friedli and Parsonage (2009)'s estimates to compute the lifetime cost of averted mental health admissions.
- In our computations, we only use the program impacts that survive adjustment of inference for multiple hypothesis testing.

References

Action for Children. 2019. "Closed doors: children's centre usage between 2014/15 and 2017/18."

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation."

Brewer, Mike, Sarah Cattan, Claire Crawford, and Birgitta Rabe. 2020. "Does more free childcare help parents work more?" Working Paper.

Cascio, E.U. 2009. "Maternal labor supply and the introduction of Kindergartens into American public schools." *The Journal of Human Resources*, 44(1): 140–170.

Child Accident Prevention Trust. 2013. "The costs of head injuries."

Conti, Gabriella, Stephen Morris, Mariya Melnychuk, and Elena Pizzo. 2017. "The economic cost of child maltreatment in the UK: a preliminary study."

Cooper, N J, D Kendrick, C Timblin, M Hayes, G Majsak-Newman, K Meteyard, A Hawkins, and B Kay. 2016. "The short-term cost of falls, poisonings and scalds occurring at home in children under 5 years old in England: multicentre longitudinal study." 22(5): 334–341.

Friedli, L., and M Parsonage. 2009. "Mental health pro- motion: Building an economic case." *All Wales Mental Health Promotion Network*.

Gaheer, S., and G. Paull. 2016. "The Value for Money of Children's Centre Services: Evaluation of Children's Centres in England (ECCE) Strand 5." Department for Education Research Report.

Gelbach, J.B. 2002. "Public schooling for young children and maternal labor supply." *American Economic Review*, 92: 307–322.

González-Izquierdo, Arturo, Jenny Woodman, Lynn Copley, Jan van der Meulen, Marian Brandon, Deborah Hodes, Fiona Lecky, and Ruth Gilbert. 2010. "Variation in recording of child maltreatment in administrative records of hospital admissions for injury in England, 1997–2009." 95(11): 918–925.

Kendrick, D, J Ablewhite, F Achana, P Benford, R Clacy, F Coffey, and et al. 2017. "Keeping Children Safe: a multicentre programme of research to increase the evidence base for preventing unintentional injuries in the home in the under-fives." *Programme grants for applied research*, 5(14).

Meadows, Pam. 2011. "National Evaluation of Sure Start Local Programmes: An Economic Perspective." Pam Meadows and the National Evaluation of Sure Start Team, Department for Education Research Report no. DFE-RR073.

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

Sammons, P., Hall J. Smees R., K. Smith T. Evangelou M. Eisenstadt N. Goff, J. with Sylva, and G. Smith. 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 Report no. RR495.

Stevens, C.A., D. Turner, C.E. Kuehni, J.M. Couriel, and M. Silverman. 2003. "The economic impact of preschool asthma and wheeze." *European Respiratory Journal*, 21(6): 1000–6.

Telford, C., C. Green, S. Logan, K. Langley, A. Thapar, and T. Ford. 2012. "Estimating the costs of ongoing care for adolescents with attention-deficit hyperactivity disorder." *Social Psychiatry and Psychiatric Epidemiology*, 48: 337–344.

Trefan, L, R Houston, G Pearson, R Edwards, P Hyde, I Maconochie, RC Parslow, and A Kemp. 2016. "Epidemiology of children with head injury: a national overview." *Archives of Disease in Childhood*, 101(6).