IFS

Institute for Fiscal Studies

Sarah Cattan
Gabriella Conti
Christine Farquharson
Rita Ginja
Maud Pecher

The health effects of universal early childhood interventions: evidence from Sure Start

Working paper

25/20

UKRI Economic and Social Research Council

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

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

May 19, 2025

Abstract

Early Childhood Interventions (ECI) offering disadvantaged children education, care and family services in the US show long-lasting health impacts. Can these benefits hold when these programs are offered in contexts with universal healthcare? We estimate the health impacts of Sure Start, a universal integrated ECI in England, from infancy to adolescence exploiting its rollout and administrative hospitalizations data. One additional Sure Start center per thousand age-eligible children increases hospitalizations by 10% at age 1, largely driven by infections. Admissions at ages 11-12 are reduced by 8-10%, and driven by injuries and mental health-related admissions. Impacts are concentrated in disadvantaged areas.
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 (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 thank NHS Digital for access to the Hospital Episode Statistics under data sharing agreement CON-305762-B8S7B. The authors are also grateful to the Institute for Social and Economic Research, the Office for National Statistics, and the Institute of Education for access to supplementary survey data. The authors gratefully acknowledge 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

With growing awareness that many health inequalities are rooted in the earliest years of life, policy-makers in both high-income and developing countries have increasingly focused attention on early years programs as a tool to improve population health and reduce inequalities. Across OECD countries, spending over the last 20 years on education and family services has grown fastest in the early years (OECD, 2019). This interest is, in no small part, due to a large, high-quality, mainly experimental evidence base from the U.S. showing that early years programs integrating education, childcare and family support services consistently deliver significant and long-lasting improvements in health.[1]

There are at least two reasons why integrated early years programs like these might be particularly powerful. 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.

But while the evidence base for these integrated programs is of very high quality, it has largely focused on programs that are highly intensive and highly targeted towards disadvantaged children. A large literature on scaling early years interventions suggests that more intensive and more targeted programs can more easily demonstrate effectiveness (see List, Suskind and Supplee (2021)

[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 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, 2023; Bailey, Sun and Timpe, 2021; Thompson, 2017). 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).

for an overview). Further, much of the existing evidence comes from the U.S., where the safety net is weaker than in many other OECD countries. This is important in part because a key channel through which programs like Perry Preschool and Abecedarian improved the health of their participants is by increasing take-up of health insurance (Carneiro and Ginja, 2014; Conti, Heckman and Pinto, 2016; Muennig et al., 2009), and many U.S. studies find that expanding access to health insurance early in life has long-term benefits for health and wider outcomes such as education and employment (Goodman-Bacon, 2021*b*; Miller and Wherry, 2019; Wherry et al., 2018).

As a result, it is unclear how effective and cost-effective integrated early years services can be when delivered universally in contexts with a stronger safety net, when they do *not* modify the availability or take-up of health insurance. Filling this evidence gap is vital for realistically assessing the potential for integrated early years services to deliver improved outcomes for children in lighter-touch, fully-scaled models. There are active debates about similar programs in many European countries as well as in the U.S., where free universal healthcare is increasingly debated.[2]

This paper addresses this gap by evaluating the health impacts of Sure Start, a universal integrated early childhood intervention in England, where the National Health Services 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 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 children's health 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. A key innovation of the paper is to estimate program impacts

[2] For example, integration of children's services has recently been 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

on different types of hospital admissions at each age, from 1 to 15, using administrative data covering 21 years of admissions to NHS hospitals (which account for 95% of all admissions). These rich estimates support a precise cost-benefit analysis and help shed crucial light on the possible mechanisms underlying the program's impacts.

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. We extensively probe the validity of our strategy by analysing the determinants of the rollout, testing for pre-trends, and showing that our results are robust to the inclusion of a range of differential trends across areas, and to heterogeneity in treatment effects across areas and cohorts (Borusyak, Jaravel and Spiess, 2024; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021*a*, among others).

Our analysis yields three main results. First, 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. For all 11- and 12-year-olds, the total was close to 5,500 prevented hospitalizations each year. We estimate the overall financial benefits of these effects to offset a third of the cost of providing the program. A placebo check for congenital chromosomal defects - which are not plausibly affected by Sure Start - finds precisely estimated null results.

Second, we show that these changes in hospitalizations reflect, at least in part, changes in underlying health. Using administrative data, we show that the average length of stay at older ages decreased, even as the total number of hospitalizations fell. This suggests that the persistent decrease in hospitalizations we find as a result of Sure Start access was not driven by less-sick adolescents selecting out of hospitalizations. Using a different, between-sibling design within a nationally representative longitudinal household survey, we also show that greater access to Sure Start improved self-reported health and mental health among adolescents.

Third, we explore the mechanisms through which Sure Start affected hospitalizations and health. This is challenging because no data linking Sure Start services to outcomes are available, but we use two pieces of evidence to answer it indirectly. 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 it may have affected: preventable conditions, infectious illnesses, accidents and injuries, and – among adolescents – mental health. We find evidence of a persistent, pervasive decline in accidents and injuries through the early years and childhood, as well as a decline in mental health-related hospitalizations in adolescence. Hospitalizations for infectious illnesses and preventable conditions increase in the early years before falling during childhood and adolescence. Second, we use another nationally representative longitudinal survey to directly test whether the program affected parental employment and family income.

Overall, our findings suggest that Sure Start's benefits arose from a range of services, including both services targeting health and those with wider aims. 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. Our evidence also suggests that access to Sure Start in the early years improved children's behavioral and emotional development through adolescence, by promoting better parenting practices and safer home environments. We rule out services focusing on increasing parental labor 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. 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 program may have played an important role in driving the improvement in physical and mental health.

These results speak 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 U.S. 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)[4] and several others focusing on policies expanding access to childcare and preschool education (Breivik, Del Bono and Riise, 2021; Hong, Dragan and Glied, 2019; Sandner, Thomsen and González, 2024; van den Berg and Siflinger, 2022).[5]

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.

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 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

[4] 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.

[5] 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. Sandner, Thomsen and González (2024) find that an increase in childcare slots by one percentage point in German counties reduced child maltreatment cases leading to out-of-home placement.

because the persistence of the effects in the post-eligibility years provides a stronger basis for predicting longer-term impacts - a key concern for 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 universal 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 impact of Sure Start. The first study, the National Evaluation of Sure Start (NESS), compared children with access to the earliest Sure Start centers against those in areas not served by the program (Melhuish et al., 2005, 2008*b*, 2010). The second study, the Evaluation of Children's Centres in England (ECCE), analyzed descriptive associations in the early 2010s between use of Sure Start services and child health (Sammons et al., 2015). Anderberg and Olympiou (2023) evaluate the impact of the later phase of Sure Start Children's Centres on children's social care outcomes, using aggregated data at the local authority level. Along with Carneiro, Cattan and Ridpath (2024),[6] who study the program's impacts on academic outcomes, our paper proposes a robust quasi-experimental design using administrative microdata, which allows us to examine the causal impacts of the program much beyond the time horizon considered by previous evaluations.

The rest of the paper is organized as follows. We provide background on Sure Start in section 2, describe our data and empirical strategy in section 3, and present results on overall hospitalizations in section 4 before exploring mechanisms and heterogeneity in section 5 and section 6 respectively.

[6] Carneiro, Cattan and Ridpath (2024) exploit the rollout of the program over time, but do not estimate the same treatment effect parameter. They exploit information on the child's postcode at age 5 to use a distance-based approach, defining the treatment variable as the number of centers within walking distance from the child's home.

2 Institutional Background: The Sure Start Programme

2.1 The rollout of Sure Start

First introduced in 1999, Sure Start was conceived as universal-within-area intervention whose services would be available to *all* families with a child under five in the neighborhood of the center (without individual means-testing). Initially, the program set up 250 'Sure Start Local Programmes' (SSLPs) in highly disadvantaged areas (Melhuish et al., 2008*a*; Pugh and Duffy, 2010); the centers' popularity meant the target was almost immediately doubled (Eisenstadt, 2011). The allocation of these initial centers was based on strict national guidelines on the local level of deprivation, the incidence of low birth weight and of teen pregnancy.

Four years later, the government announced that Sure Start would transition from a program offering universal services within disadvantaged neighborhoods into a fully universalized offer, with a target of 3,500 centers to deliver "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 covering outreach, health services, and links to childcare and employment support (Lewis, 2011).

This stage of the rollout was intended to be driven entirely by local deprivation, with three well-defined phases intensifying provision in the 20% most deprived neighborhoods; expanding the offer to the 30% most deprived neighborhoods; and univeralizing access (House of Commons, 2010). Even after the full rollout, disadvantaged areas retained more intensive provision: 70% of the centers planned to be in the 30% most disadvantaged areas.

By 2010, the overall number of centers reached nearly 3,300, 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 trend in number of centers and program budget). At its peak in 2010, Sure Start accounted for about a third of overall spending on programs for the under-5s in England, attesting to the national importance of the policy (Britton, Farquharson and Sibieta, 2019). Despite the rapid rollout, Figure A.2 suggests that the guidelines were largely adhered to, with three distinct rollout phases by deprivation.

The new government elected in 2010 de-prioritized Sure Start and, between 2011 and 2019, spending on the program fell by over 60% (Britton, Farquharson and Sibieta, 2019). Following the removal of earmarked funding in April 2011, local authorities responded in different ways: some subsidized Sure Start services from other budget lines, while others consolidated several centers into one, cut back on services or (more rarely) shut down centers entirely (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.

2.2 Services offered by Sure Start

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

As with Head Start in the U.S., the mix of services offered by Sure Start varied across areas, reflecting local priorities. In general, the services offered by Sure Start lay in four main domains: health, parenting support, childcare, and parental job assistance.

Health services supplemented existing services available through the public healthcare system (for example, running drop-in sessions with a midwife or health visitor) and provided additional support that was not routinely available (such as breastfeeding support; advice on accident and injury prevention; advice on diet and nutrition; and support for parental mental health). The centers also had strong links with the local public healthcare system and could provide advice about looking after children as well as signposting to healthcare services where necessary (e.g. DfES, 2003).

Sure Start centers also offered **parenting support**. These services ranged from informal drop-in sessions which facilitated contact between parents, to more intensive interventions such as evidence-based parenting programs like Triple P or Incredible Years. While a minority of children accessed **childcare** directly through Sure Start, centers were required to refer parents to other settings where they could take up their government-funded part-time childcare place (see below). Finally, centers offered **adult support services** such as training for volunteers, basic skills classes, advice on welfare and housing, and links to employment services.

2.3 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. The integration of services into one hub may have increased the take-up of existing services.

Second, Sure Start **provided new programs to meet local unmet demand** (DfEE, 1999). Among those, Sure Start provided health services, such as breastfeeding support, 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 sessions, for free or a very low price.

Third, Sure Start may have **changed families' childcare use**, both directly through provision of care and indirectly through referrals. Data from 2011 showed that around 15% of 3- and 4-year-olds used childcare provided by Sure Start (Goff and Chu, 2013), meaning that Sure Start-based childcare directly affected only a minority of children. However, Sure Start centers were also required to signpost families to childcare providers where they could take up their 'free entitlement' to a part-time funded childcare place.[7] Moreover, Sure Start may have affected the quality of the childcare used, both directly (via provision of on-site care) and indirectly (via referrals to higher-quality providers).

[7]Introduced in 1997, the free entitlement program was progressively expanded during the 2000s. By 2010, all 3- and 4-year-olds in England were eligible to receive 15 hours a week of funded childcare for 38 weeks a year. Later in the paper we will confirm that our results identify the impact of Sure Start itself, rather than the contemporaneous rollout of this childcare program.

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

In the 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.

In the early phase of Sure Start's rollout (2000-02), 28% of families report knowing of Sure Start and 5% of them report having used Sure Start. As we show in Appendix Table A.2, data from the nationally-representative Millennium Cohort Study suggest that both awareness and usage are significantly higher in Local Authorities with a greater concentration of Sure Start centers. Controlling for child and family characteristics, we find that one extra center per 1000 children age 0-4 is associated with a 13 percentage point increase in the probability of using Sure Start.

Our second data source, from the early 2010s, follows parents who were registered with a Phase 1 or Phase 2 center (Goff and Chu, 2013), located in the 30% most deprived areas of the country. By this point, almost all families were automatically registered with a center from the birth of their child. Overall, around 60% of registered families used at least one Sure Start service, rising to 75% of those with a 1-year-old. Figure 1 shows that, across all ages, the most commonly used services were health services and parenting support, each reaching around 40% of children. By contrast, around 14% of 3- and 4-year-olds used Sure Start childcare (younger children were not eligible for free childcare at this point).

Appendix Figure A.3 shows that families spent on average 2-3 hours a week using Sure Start services, with the most intensive use of parent-child services like drop-in play sessions. While higher-income families spent more time using health services, lower-income families made heavier use of parenting support and adult support services. This is confirmed in Appendix Table A.3, which shows that Sure Start use among the youngest children was higher for those from less disadvantaged backgrounds, while take-up among two- to four-year-olds was higher among more disadvantaged families.

3 Data and empirical strategy

3.1 Data sources

Our main impact analysis uses administrative data on hospitalizations, combined with data on the precise location and opening date of Sure Start centers.

Data on Sure Start facilities Our treatment variable measures local Sure Start coverage during the period when a child was eligible to attend. To construct this, we use a unique dataset with the exact address and date of opening of each Sure Start center between 1999 and 2010.[8]

Our coverage measure varies across Local Authority (LA) d and quarter of birth q (our cohort dimension):[9] specifically, we define SS_{dq} as the average over the first 60 months of life of the number of centers per thousand children aged 0-4 in the child's LA.[10] 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 the treatment variable for each of the 323 LAs in England (in gray) and superimposes its average (in blue) across LAs. Overall, the median LA had coverage of 0.14 centers per thousand children; however, this includes many untreated cells. By the time the program had been fully rolled out in 2010, the median LA operated just under one center per thousand children.

Data on hospitalizations Our primary outcomes are (all-cause and cause-specific) hospital inpatient admissions. This includes all patients admitted to a hospital bed (even if not kept overnight). To measure these, we use administrative data from the Hospital Episode Statistics (HES) on all

[8] Our treatment variable does not distinguish between Local Programmes and Children's Centres. In line with historical evidence (Hands et al., 2006), we assume that all Local Programmes had transitioned into Children's Centres by December 2006, unless we observe an earlier transition.

[9] 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.

[10] There are 326 Local Authorities (LAs) in England. We exclude three very small LAs, which are outliers in terms of coverage, from the analysis (the Isles of Sicilly, City of London, and West Somerset).

admissions to English public hospitals between April 1997 and March 2018. HES holds information on admission and discharge dates and clinical diagnoses recorded via ICD-10 codes[11] linked to demographic information on the patient's sex, ethnicity, date of birth, and neighborhood (or Lower-level Super Output Area) of residence at the time of admission.[12] 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 hospital admission, excluding admissions related to the birth of a child. We then construct counts of all-cause andcause-specific admissions for each neighborhood (defined at the LSOA level by quarter of birth by sex by age of admission cell. Cells without admissions are assigned zero. Because close to 90% of cells have zero admissions (and just 3% of cells have multiple all-cause 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).

3.2 Econometric specification

Our aim is to estimate the effect of increased access to Sure Start on children's hospitalizations. We exploit the variation in SS_{dq}, which measures the potential exposure to Sure Start across birth cohorts and Local Authority generated by the Sure Start rollout (displayed in Figure 2). We operationalize this in a standard difference-in-differences framework by way of a two-way fixed effects model, where we control for both birth cohort fixed effects (to account for secular trends in hospitalization) and neighborhood (LSOA) fixed effects (to account for systematic differences in time-invariant area characteristics). Importantly, these LSOA fixed effects will largely control for area-level differences in deprivation, which drove the rollout of Sure Start and is highly correlated

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

[12] The Lower-level Super Output Area (LSOA) is a very small geographic unit which nests within a Local Authority. Each LSOA has, on average, a population of 1,500 individuals. There are about 33,000 LSOAs in England.

with hospitalizations.[13]

Our main estimating equation has the following specification:

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

where $D^{ya}_{sql(d)}$ and $SS_{d(q)}$ are defined as above. X_s is an indicator taking the value 1 if $D^{ya}_{sql(d)}$ refers to neighbourhood-birth quarter-level outcome for females and 0 otherwise. We control for $Pop_{al(d)}$, the number of children of age a in neighborhood l, because the probability of observing non-zero hospital admissions rises with population size.[14] γ^{ya}_q is a set of quarter-year of birth fixed effects. The model includes a set of over 32,000 neighborhood (LSOA) fixed effects $\pi^{ya}_{l(d)}$, which account for time-invariant unobserved heterogeneity across areas. Finally, the error term is denoted $\upsilon^{ya}_{sql(d)}$; 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).

In this specification, 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. In order for the parameter δ^{ya} to recover the causal impact of greater access to Sure Start on children's hospitalizations, we require several conditions to hold: (1) that greater access to Sure Start increases the probability of participation; (2) that families did not locate selectively to be closer to Sure Start centers as they rolled out; and (3) the "parallel trends" assumption that the rollout of Sure Start across LAs was uncorrelated with time-varying unobservable determinants of or shocks to hospitalizations (captured in $\upsilon^{ya}_{sql(d)}$).

Our discussion of the take-up of Sure Start services in subsection 2.4 provides descriptive evidence that greater access to Sure Start was related to greater use of its services, thus validating (1).

---

[13]Deprivation, as measured by the Index of Multiple Deprivation, is highly persistent over time. For example, the correlation between LSOAs' rank in the 2000 and 2004 Index of Multiple Deprivation is 84%; the correlation between 2004 and 2010 is 97%.

[14]By definition, population size is relatively similar across LSOAs. However, at the level of single year of age there are some differences both across LSOAs and within LSOAs over time, so we include population controls in our preferred specification. We show in Table A.4 that our results are robust to weighting cells by population, while excluding these population controls generates larger estimates.

In Appendix B, we use data from a representative longitudinal household survey to explore patterns of migration in families with young children. We find relatively little migration across LAs overall (with around 4% of families moving LA each year after children turn 5) and no relationship between Sure Start coverage and migration in the early years. This provides reassurance for criterion (2), that selective migration is not a major concern in this setting. Finally, in subsection 4.2 we probe the plausibility of assumption (3) on the "parallel trends" in hospitalizations.

4 Sure Start's effects on overall hospitalizations

4.1 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). As we study a relatively large number of outcomes, we also report the results of a stepwise multiple hypothesis testing procedure that controls for family wise error rate (Romano and Wolf, 2005).[15] To enable comparison of relative effects across ages, Figure 3 plots these estimates re-scaled by the baseline probability (measured on the cohort born in 1996) of any hospitalization at the corresponding age.

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. This translates into about 6,700 more yearly hospitalizations.

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

[15]Specifically, we use the procedure in algorithms 4.1 and 4.2 of Romano and Wolf (2005) to account for testing several hypotheses simultaneously. We use 500 block-bootstrap replications to obtain the adjusted critical values (the block is the LA). 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).

children turn 5 and stop being age-eligible to use Sure Start services, the overall impact on hospitalizations becomes consistently negative, with significant decreases in hospitalizations at ages 11 and 12. 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 that the increase in admissions among infants and the reductions at ages 11 and 12 survive the adjustment of inference to multiple hypothesis testing.

In section 5, we will show that the effects at different ages are driven by different causes of hospitalization: a substantial increase in infectious illness and hospitalizations for preventable causes at the youngest ages, a decline in mental health-related admissions in adolescence, and a persistent and pervasive decrease in accidents and injuries throughout the early years and childhood.

4.2 Validity of the empirical strategy

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 assess the plausibility of this assumption. We use "context-specific economic knowledge" (Roth, 2022) coupled with a range of empirical tests to assess whether there are time-varying factors correlated with both the rollout of Sure Start and hospitalizations that could bias our empirical strategy. We also check empirically whether there were "parallel trends" in hospitalizations in the pre-Sure Start period. We also conduct placebo checks, including showing a precisely estimated zero impact of Sure Start on genetic conditions.

4.2.1 Context for the rollout of Sure Start

As outlined in section 2, the Sure Start rollout was driven by a clear policy of prioritizing areas based on a small number of observable characteristics, chief among them local neighborhood deprivation.[16] Since England's official measure of neighborhood deprivation is highly persistent (see Footnote 13), controls for neighborhood-level fixed effects should capture most of the systematic differences between areas that were more and less heavily exposed to Sure Start.

Of course, policymakers could have been influenced by other factors in deciding when or where to open the next center. Our identification strategy would be under threat if these unknown factors driving the rollout varied across areas and over time, and also themselves affected children's hospitalizations.

In practice, we find that area and cohort fixed effects explain 86% of the variation in SS coverage. We construct a panel dataset with a wide range of variables capturing the local context, including local population changes; the provision of related services like GPs or childcare places; local labour market characteristics; and local political alignment with the national government (see Appendix C for the full description of variables used). This set of local characteristics explains only 4% of the variation in the rollout, suggesting that the rollout was indeed mostly determined by time-invariant area characteristics.

We build on this analysis by testing how the year-to-year, within-LA variation in SS exposure correlates with a large set of baseline area characteristics, measured just before Sure Start was rolled out.[17] We consider both the stated 'guideline' variables (deprivation, teen pregnancy and low birth weight) as well as an extensive set of additional time-varying local characteristics. As Appendix Figure A.4 shows, while the expansion of Sure Start coverage was significantly associated with guideline variables, we find no evidence that other characteristics systematically predicted the rollout of Sure Start. The one partial exception is the local unemployment rate (measured by claimants of unemployment benefits), which is positively and significantly correlated with the expansion of SS coverage in the early part of the rollout. While this may be unsurprising given that Sure Start aimed to improve parental employment, we exert caution and include local unemploy-

[16] At the start of the rollout, policymakers additionally used information on low birth weight and teen pregnancy rates from around 1998 to guide the location of the earliest Local Programmes.

[17] We do using the same specification as Bhuller et al. (2013) whereby we estimate the following equation: $\Delta SS_{dq} = \rho_d + [\beta_q \times c_{d,1998}]' \Phi_q + \epsilon_{dq}$ 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.4.

ment alongside guideline variables in further robustness checks below.

Finally, to further check whether unobservable time-varying factors likely influence both the rollout of Sure Start and the hospitalization rate, we following Roth et al. (2023) by re-estimating our benchmark specification (Equation 1), controlling for 11 time-varying local authority characteristics (including the local unemployment rate). We run two specifications. The first, shown in the blue line in Figure A.5, measures these controls in the birth year of the corresponding cohort.[18] The second, shown in the red line of the same graph, measures the controls in the admission year. This latter specification tests whether our main results are driven by a correlation between Sure Start's rollout and later time-varying characteristics such as service provision (e.g. if policymakers sought to 'follow up' early intervention with later services). In both cases, the inclusion of the controls make very little difference to our results, with no statistically significant difference to our benchmark results (shown in green). 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.

4.2.2 Pre-treatment trends in hospitalizations

Our identification strategy would also be compromised if the rollout of Sure Start were directly correlated with (potential) hospital admissions. Although this is not directly testable, we provide evidence that this is unlikely by testing whether the rollout of Sure Start was correlated with pre-existing trends in hospitalizations. To do so, we we restrict our sample to neighborhood-cohorts before treatment onset, and estimate:

$$D^{ya}_{sql(d)} = \sum_{l=2}^{l=L} \delta^{ya}_l SS_{d,q+l} + \beta^{ya} X_s + \alpha^{ya} Pop_{al} + \gamma^{ya}_q + \pi^{ya}_{l(d)} + \upsilon^{ya}_{sql(d)} \quad (2)$$

[18] 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. We interpret the robustness of our results to directly controlling for the rollout of free childcare places as evidence that our main estimates are picking up the effects of Sure Start, rather than childcare. For additional reassurance, we also estimate a model with only childcare take-up as a time-varying control, finding similar results and generally insignificant coefficients on the free entitlement variable (results available on request).

where L is the maximum number of pre-trend coefficients we estimate in the regression, and the reference category is an average of cohorts $l = 1$ and $l - 2$ (following (Borusyak, Jaravel and Spiess, 2024), we omit two leads to avoid under-identification). As recommended in Schmidheiny and Siegloch (2023), we experiment with a number of values for L, ranging from 24 to 40 quarters pre-treatment, trading off length of pre-trends against sample size, with qualitatively similar results for each. In Figure 4, we report estimates based on L=24 as it allows us to use a common number of leads across all age groups.

For each age of admission, we report the p-value of a joint test of the pre-trend coefficients. While the null hypothesis that these coefficients are jointly zero is rejected at the 5% significant level for a few ages (2, 3, 13, 15), inspecting the coefficients visually suggests that there is little evidence of a clear pre-trend in the direction of treatment even at these ages.[19]

As an additional check for pre-trends, we assess the robustness of our results to allowing for differential trends across areas. We specify these different trends in two ways. First, we augment model Equation 1 with interactions between a cohort trend and the baseline characteristics that we showed persistently affected the rollout of Sure Start (1998 deprivation levels, teen conception rate, incidence of low birth weight, and local unemployment). Table A.5 shows that our main estimates remain largely unchanged in these specifications, whether we use a linear or quadratic cohort trend.

Second, we allow for trends in hospitalizations to vary across individual local authorities. To avoid incorporating the effect of Sure Start into these trends, we first estimate LA-specific linear time trends in the pre-treatment period (when $SS_{dq} = 0$), then linearly extrapolate this LA-specific pre-treatment time trend for all cohorts in the sample. We include this estimated trend as a control in our main model (Equation 1). These estimates are presented in Appendix Figure A.6 and are

[19]Note that in this specification, we cannot use our main definition of Sure Start treatment as the average coverage experienced in the first 60 months of life, since averaging treatment in this way introduces a strong degree of collinearity in the *change* in treatment intensity from one period to the next (which in turn introduces a high degree of collinearity in the leads and lags of change in treatment status used in the continuous time-to-event specification). Instead, we define $SS_{d,q+l}$ as the level of Sure Start coverage in the last relevant month of exposure. For outcomes measured at age 5 onwards, we use treatment at month of age 59. For earlier outcomes, we use treatment one month before age A for outcomes defined at age A.

similar to our main estimates of Figure 3.

4.2.3 Placebo tests

Finally, we conduct a placebo analysis, 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 – and find – no impact of Sure Start on admissions for these conditions, with precisely estimated null effects (Table 2).

4.3 Treatment effect extensions

Our main estimates focus on estimating the profile of the effect of Sure Start coverage on hospitalizations at different ages. For simplicity, we have so far focused on estimating a single treatment effect for each age. In this section, we explore a range of extensions to this analysis.

Intensive margin In Table A.6, we complement our main extensive-margin results with two additional outcomes reflecting the intensive margin: the *total* number of hospitalisations in each cell (Panel A), and the average length of stay (Panel B). Panel A shows a similar, though less precisely estimated, age profile to our main results. In Panel B, we find no significant impacts on length of stay at younger ages, but some decreases later on. We return to discuss these in section 5.

Non-linear effects In addition, we can explore whether there are non-linear impacts from Sure Start coverage, for example if effects arise only once families have access to several centers in their vicinity. In Appendix Figure A.7, we distinguish between no Sure Start coverage, medium coverage (fewer than 0.25 centers per thousand children - the median among those with positive coverage), and high coverage (more than 0.25 centers per thousand). In line with a linear specification, we find that the impacts of high coverage are consistently of greater (absolute) magnitude than the impacts of medium coverage, though the confidence intervals are too large to reject they are statistically different from each other.

Dynamic effects Our benchmark specification estimates a single treatment parameter, averaging the effect of exposure to Sure Start over the relevant period outcomes. We can also explore the dynamics of the treatment effect, testing whether the impact of Sure Start varies over time after treatment begins. We operationalize this with an event study model for a continuous treatment variable, following Schmidheiny and Siegloch (2023). Specifically, for each age $a = (0, \ldots, 15)$, we estimate:

$$D^{ya}_{sql(d)} = \sum_{l=\underline{L}}^{\bar{L}} \delta^{ya}_{l} SS_{d,l} + \beta^{ya} X_s + \alpha^{ya} Pop_{al} + \gamma^{ya}_{q} + \pi^{ya}_{l(d)} + \upsilon^{ya}_{sql(d)} \tag{3}$$

where $\underline{L}$ and $\bar{L}$ are the maximum number of leads and lags that define the effect window (with effects outside these endpoints binned, as recommended in Schmidheiny and Siegloch (2023)).[20]

The estimates of these specifications are shown in Appendix Figure A.8. Due to the high number of coefficients, we also estimate this specification on yearly data (collapsing the dataset from quarter-year of birth to year of birth) to give a smoother profile of dynamic effects. These results are shown in Appendix Figure A.9. In both specifications, there are positive and growing effects of exposure to Sure Start in earlier ages and negative and growing effects of exposure to Sure Start at later ages, in line with our benchmark specification.

Heterogeneous treatment effects A recent literature has highlighted some limitations with using the traditional two-way fixed effects (TWFE) estimator employed here in contexts where treatment onset is staggered and effects may be heterogeneous (Borusyak, Jaravel and Spiess, 2024; Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021*a*). This is pertinent to our context: treatment onset is staggered, and heterogeneous treatment effects are

[20] We focus on an effect window of 24 quarters (6 years) after Sure Start treatment onset, for both conceptual and practical reasons. Conceptually, since we use treatment measured just before a child's fifth birthday, only cohorts born more than five years after treatment onset will be 'fully exposed' from birth. A 24-quarter window allows us to capture effects throughout this window and a little beyond. Practically, restricting to 24 quarters also means that we can estimate the same horizon of effects for each treatment age; including substantially more lags is not possible for older outcome ages. We have also estimated results for up to 40 quarters post-treatment (for the age groups where this is possible), with qualitatively similar effects; those results are available on request.

plausible.[21]

Unfortunately, no heterogeneity-robust estimator has yet been developed for a continuous, time-varying *and* staggered treatment as we use in our main results.[22] We therefore apply the efficient imputation estimator developed by (Borusyak, Jaravel and Spiess, 2024) (BJS henceforth) to discrete measures of Sure Start treatment, comparing the results to the traditional TWFE estimator in each case.

We consider three different binary treatment variables, defined as indicators for whether SS_{dq} is above 0, 0.1 and 0.25. In choosing the cut-off point for binary treatment, there is a trade-off between ensuring that the treatment group experiences a meaningful dose, while minimizing the share of the control group that is also meaningfully treated.[23] To strike a balance between these considerations, our preferred cut-off is therefore 0.1 centers per 1,000 children in the local authority.[24]

The results comparing these binary treatment effects are in Table A.8. In general, the TWFE and the BJS estimates are very similar, particularly for $SS_{dq} > 0.1$ and $SS_{dq} > 0.25$. There are some greater differences if binary treatment is defined by $SS_{dq} > 0$, though we note that the TWFE estimates here also look markedly less similar to our main specification (suggesting that, indeed, the cut-off is too low to effectively pick up variation in the intensity of Sure Start treatment). Overall, this suggests that negative weights are unlikely to be driving our main results.

[21]Following the diagnostic test proposed by de Chaisemartin and D'Haultfœuille (2020), we estimate that, across all our models, the proportion of negative weights is around 50%. Their sums vary between -0.17 and -0.53 (see Table A.7)

[22]Callaway, Goodman-Bacon and Sant'Anna (2024) propose an estimator for continuous treatment, but require that "the amount of the treatment remains constant in post-treatment periods." de Chaisemartin, D'Haultfœuille and Vazquez-Bare (2024)'s estimator for continuous treatment relies on a sufficient population of "stayers" in each period who do not switch treatment. Our set-up does not meet either of these requirements.

[23]Simply dropping observations with positive treatment but below the chosen cut-point is not possible with the newer difference-in-differences estimators, which require a strongly balanced panel.

[24]With a cut-off at 0.1 center per thousand, more than 70% of local authorities switch straight from zero coverage to 'treatment' - meaning that this measure has a fairly clean control group. By contrast, under the higher cutoff of 0.25 centers per 1,000 children, more than 80% of local authorities would experience substantial levels of Sure Start exposure in the 'control' period. This makes it more difficult to carry out standard tests for pretrends, since our 'control' group is, in fact, exposed to Sure Start. On the other hand, a cut-off of zero may be too low, especially given evidence of a dose-response relationship in Figure A.7.

5 Mechanisms

The integrated mix of services offered by Sure Start centers and the lack of data linking services to outcomes makes it more challenging to disentangle specific channels through which the program improves child development. We shed light on this crucial question in two ways. First, we make hypotheses about the impacts that different channels would have on hospital admissions for different causes and test those using HES. Second, we analyze survey data with direct measures of physical and mental health to distinguish between improvements in health and changes in healthcare use.

5.1 Expected effects of Sure Start on health and hospitalizations

Table 3 summarizes the overall expected effects of a multi-faceted program like Sure Start on different types of hospitalizations. We describe our key hypotheses below.

Health services Sure Start centers offered a range of health services, information, and advice, which augmented but did not replace primary care provision (which in England is delivered free at the point of use). These supplemental health services may have affected hospitalizations through two main channels. First, **screening** children for conditions and referring families to appropriate healthcare could lead to an increase in hospitalizations for preventable and manageable conditions in the short term (early years) and a decrease in hospitalizations for the same conditions in the longer term. Second, **enhancing children's healthy environments** through health-promoting parental behavior and improving the safety of the home environment could reduce hospitalizations at all ages. Information on safe home environments (such as 'child-proofing' and accident and poisoning prevention) may have been particularly important during the early years and childhood.[25]

[25]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 age 5. We found no evidence of effects of the program on obesity, and so rule out this mechanism going forward.

Parenting support and parent-child services Centers provided evidence-based parenting programs to improve family functioning and positive parenting skills, often with a particular focus on children's social and emotional development.[26] They also provided a range of lighter-touch parent-child activities, such as drop-in play sessions, aimed at strengthening parent-child relationships.

While these services did not target (physical) health specifically, they may have had indirect benefits by activating cross-productivities between behavioral and emotional development and health (Cunha et al., 2006). 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** (Hoare and Beattie, 2003).

Therefore, 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 (though these are only routinely diagnosed in hospitalizations from adolescence).

Childcare and group-based sessions Higher take-up of (higher-quality) childcare could have affected health in two ways. First, high-quality childcare can **benefit emotional and behavioral development** (Heckman, Pinto and Savelyev, 2013), with similar impacts to those from parenting support. Second, childcare - as well as the other group-based activities provided by Sure Start centers - increased children's contact with others 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 (Daysal et al., 2024; Henderson et al., 1979; van den Berg and Siflinger, 2022). In this case, we would expect hospitalizations for infections to increase in the

[26] Examples of evidence-based parenting interventions delivered by Sure Start centres are Triple P (PPP) and Incredible Years (IY), which both have substantial evidence showing improvements in child behavior and socio-emotional development (Morpeth et al., 2017; Parenting and Centre, 2025).

short term and drop in the medium term, at least until a few years after all children start school at age 5.

Adult support The last major set of services offered by Sure Start supported parents, especially in their effort to gain employment. Higher parental employment could allow parents to **buy more and/or higher quality inputs**, such as more nutritious food (Carneiro and Ginja, 2016). However, parents shifting into employment could also **reduce time spent on health-improving activities** (e.g. cooking a home-made meal or accompanying children to the doctor) and increase time in childcare (e.g. Løken, Lommerud and Holm Reiso, 2018). Since these channels push in different directions, the overall effect of Sure Start's employment services on children's health and hospitalizations is ambiguous. Since this channel does not lead to clear testable predictions in the hospitalization data, we present a separate estimation of the effect of Sure Start on parental employment using data from the Labour Force Survey (Appendix D).

5.2 Service provision and cause–specific hospitalizations

Cause-specific hospitalizations As Table 3 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. We now present estimates of impacts on cause-specific hospitalizations at different ages to infer what mechanisms were most likely at play.

We consider the four types of conditions mentioned above: preventable conditions, infectious illnesses, external causes (accidents, injuries and poisonings), and (among adolescents) mental health. 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).[27] As before, we present results graphically in Figure 5 and report estimates and p-values adjusted for multiple hypothesis testing in Appendix Table A.9.

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

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, the increase fades out by the time that children start school, with some evidence of reductions in hospitalizations for these conditions thereafter. This pattern is consistent with Sure Start centres identifying health problems in early life and providing information and signposting to help parents learn how to manage their child's conditions.

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; however, there are 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. These results are consistent with increased exposure to pathogens during group-based sessions at Sure Start centers, including programs like 'stay and play', baby massage or group parenting support, which were among the services parents and children attended most on average (Sylva and Sammons, 2015). 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'.

Next, we turn to hospitalizations for external causes.[28] Figure 5(c) shows a very large decline in hospitalizations for external causes that persists throughout the early years and childhood. At the youngest ages, the probability of an externally caused hospitalization falls by 10% or more with greater access to Sure Start. Looking within the subcategories of external hospitalizations, Table A.10 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.[29]

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

[29]Injuries (ICD-10 groups V and Y) account for between 70 and 80% of external admissions; most of the rest are

The magnitude and persistence of the effect on injuries is 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). 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).[30]

Lastly, we study the impact of Sure Start on children's mental health-related admissions. Hospitalizations before adolescence are not commonly flagged with a mental health diagnosis code; we therefore focus on hospitalizations from age 12 onward. In Figure 5(d) (and Appendix Table A.11), we show that additional access to Sure Start significantly decreases mental health-related admissions at ages 12 to 14.[31] These estimates should be interpreted with caution: not only is the prevalence of mental health hospitalizations low even in adolescents, but a hospitalization-based measure will miss the majority of young people who are receiving services in the community, through their schools or 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 5.3, we complement the results presented here with analysis of survey data directly estimating the impact of Sure Start on self-reported measures of socio-emotional development among adolescents. Taken together, we interpret these results as another piece of evidence suggesting that Sure Start improved the socio-emotional development of children.

Parental employment Unlike the other channels we have discussed, parental employment effects did not have clear testable predictions in the hospitalization data. We therefore use survey data from the UK Labour Force Survey (LFS) to present direct evidence of the impact of greater

accounted for by poisonings (codes T15-T98).

[30] 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 (González-Izquierdo et al., 2010), but the incidence of these is too low to reliably estimate Sure Start's impacts on these outcomes directly (results available upon request).

[31] The most common mental health conditions among hospital admissions for these ages are acute alcohol intoxication (ICD10 code F10), which accounts for 27% of such admissions.

access to Sure Start on maternal labor market outcomes. Appendix D 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).

5.3 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. Given the nature of its services, Sure Start could have affected both margins. In this section, we provide evidence that Sure Start did change underlying health, not only healthcare usage.

Proxies in administrative data Our first approach is to examine the impacts of Sure Start on additional proxies for health captured in hospitalization data. Using our benchmark specification in the hospital admissions data, we examine the impacts of greater access to Sure Start on emergency and elective admissions to the hospital. Figure 6 clearly shows that Sure Start's overall impact on hospitalizations comes almost exclusively from changes to emergency admissions, with null effects on elective admissions for most ages. This strongly suggests that fewer children are experiencing health emergencies that lead them to be admitted on short notice via the emergency room.

The results on length of stay in Table A.6 also indicate an improvement in underlying health. At younger ages, hospitalizations increased, but we see no significant impact on the average length of stay (suggesting that the pool of hospitalized children did not become, on average, less healthy as it grew). At older ages, when overall hospitalizations decrease, we find that the average length of stay also decreases, suggesting that improvements in young people's overall health outweighed the negative selection effects for the hospitalized group.

Direct measures of health from survey data We complement our results on hospitalizations with analysis of self-reported health measures in the UK Household Longitudinal Study (UKHLS).

UKHLS is an annual longitudinal panel of a representative sample of around 40,000 households 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).

Our main specification, which relies on 33,000 LSOA-level fixed effects, is too demanding to implement in survey data. Instead, we leverage the longitudinal nature of 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 we estimate).

As reported in Table 4, 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 decline in hospitalizations among adolescents is related to improvements in underlying physical and mental health. It also serves as an important check on our primary results, showing corroborating evidence based on a totally different empirical strategy.

6 Impact heterogeneity by gender and deprivation

The literature evaluating early childhood interventions often reports that interventions are differentially effective for 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.

6.1 Heterogeneity by gender

Evaluations of early childhood interventions commonly find larger effects for boys (see for example the Abecedarian program (Conti, Heckman and Pinto, 2016), Head Start (Carneiro and Ginja, 2014) and the Boston Preschool program (Gray-Lobe, Pathak and Walters, 2022)). We examine whether this is true for Sure Start.

Figure 7 (and the point estimates in Table A.12) 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. An analysis of gender-specific effects on hospitalizations for different causes reveals that the gender difference in the program's impacts during adolescence is entirely driven by the greater impact of Sure Start reducing hospitalizations for injuries for boys (see Figure A.10).

In Table A.3, we show that the only gender differences in Sure Start take-up were for parenting services, which we expect to mainly influence hospitalizations through the home environment and children's socio-emotional skills. This suggests that stronger impacts for boys could result from a higher level of take-up of the services with the strongest long-run effects. Another (not mutually exclusive) possible explanation is that there is more scope for boys to benefit from the common service offer over the outcome horizon we consider; for example, Bertrand and Pan (2013) note that behavioral problems tend to start earlier for boys, so there may be more scope for Sure Start to reduce hospitalizations in this group.

6.2 Heterogeneity by level of deprivation

Sure Start was launched 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). We now assess whether this is also true of Sure Start.

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 (or LSOA) of residence.[32] In particular, we allow for heterogeneity of Sure Start effects for three groups of neighborhoods: the 30% most deprived neighborhoods, the 30% least deprived neighborhoods, and those in the middle of the distribution of disadvantage. As Figure 8 and Table A.13 illustrate, the profile of results 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 age 7 onwards, effects for children in the 30% least deprived neighborhoods are precisely estimated to be zero, while children in the poorest 30% of neighborhoods see up to a 10% decrease in the probability of hospitalization.[33]

The stronger impacts of Sure Start among more deprived neighborhoods could be a result of a number of different and non-mutually exclusive factors, including differential counterfactual environments, differential service quality, and and differential service take-up. First, there is much evidence to show that disadvantaged children grow up in less safe and stimulating environments and that their parents make less use of healthcare (Currie, 2006). They could, as a result, have have had more scope to benefit from the information and services to support parents that Sure Start provided. Second, 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, for example with more stringent requirements to provide services directly (rather than through signposting). Finally, impacts of accessing Sure Start in more deprived areas may be higher than in less deprived areas because underlying service take up was higher, a pattern confirmed by the survey data on service take shown in Figure A.3.

[32]To classify neighborhoods into these three groups, we use the 2004 Index of Multiple Deprivation (IMD), which is the government's official measure of small area deprivation (and which formed the basis of the deprivation-driven rollout of Sure Start Children's Centres).

[33]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 2018, 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.11, which shows that there is still a lot of variation in exposure to Sure Start for relevant cohorts in the middle and richest neighborhoods.

7 Conclusion

Early childhood interventions that offer disadvantaged children preschool education and family support services in the U.S. 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 current policy efforts to integrate and expand early years services in a number of OECD countries.

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. This paper presents the first robust quasi-experimental evaluation of the health benefits of accessing Sure Start centers.

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 10.7% at age 12. These results were driven by effects in poorer neighborhoods and - during adolescence - on boys.

Exploring Sure Start's impact on different types of hospitalization, we find a persistent and pervasive decrease in accidents and injuries. This is offset in the early years by an increase in infectious illnesses. During adolescence, we find a substantial decrease in hospitalizations for mental health conditions. We confirm these results with a complementary sibling fixed effects analysis in survey data, which shows substantial improvements in self-reported physical and mental health among young people who had greater early exposure to Sure Start.

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; providing better information about or access to safe environments; strengthening children's immune systems; and improving children's behavioral and emotional development, by improving parenting practices and/or providing high-quality childcare. 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.

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 F). 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 (Anderberg and Olympiou, 2023; Carneiro, Cattan and Ridpath, 2024). 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).

Anderberg, Dan, and Christina Olympiou. 2023. "Children's social care and early intervention policy: Evidence from sure start." *Economica*, 90(359): 953–977.

Anders, John, Andrew C. Barr, and Alexander A. Smith. 2023. "The Effect of Early Childhood Education on Adult Criminality: Evidence from the 1960s through 1990s." *American Economic Journal: Economic Policy*, 15(1): 37–69.

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 J., Shuqiao Sun, and Brenden Timpe. 2021. "Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency." *American Economic Review*, 111(12): 3963–4001.

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 R. Gibbs. 2022. "Breaking the Cycle? Intergenerational Effects of an Antipoverty Program in Early Childhood." *Journal of Political Economy*, 130(12): 3253–3285.

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.

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

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies*, 91(6): 3253–3285.

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. 2024. "Difference-in-Differences with a Continuous Treatment." National Bureau of Economic Research 32117.

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science*, 343(6178): 1478–1485.

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): 135–73.

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

Carneiro, Pedro, Sarah Cattan, and Nick Ridpath. 2024. "The short- and medium-term impacts of Sure Start on educational outcomes." Institute for Fiscal Studies.

Cattan, Sarah, Gabriella Conti, Christine Farquharson, and Rita Ginja. 2019. "The health effects of Sure Start." Institute for Fiscal Studies.

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, 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, Chapter 10, 266–290. New York, NY:Russell Sage Foundation Publications.

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.

Daysal, N. Meltem, Hui Ding, Maya Rossin-Slater, and Hannes Schwandt. 2024. "Germs in the family: The short- and long-term consequences of intra-household disease spread." National Bureau of Economic Research Working Paper 29524.

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

de Chaisemartin, Clément, Xavier D'Haultfœuille, and Gonzalo Vazquez-Bare. 2024. "Difference-in-Difference Estimators with Continuous Treatments and No Stayers." *AEA Papers and Proceedings*, 114: 610–13.

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." U.S. Department of Health and Human Services.

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

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.

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." *Archives of Disease in Childhood*, 95(11): 918–925.

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

Goodman-Bacon, Andrew. 2021*b*. "The long-run effects of childhood health insurance coverage: Medicaid implementation, adult health and labor market outcomes." *American Economic Review*, 2011: 2550–2593.

Gray-Lobe, Guthrie, Parag A Pathak, and Christopher R Walters. 2022. "The Long-Term Effects of Universal Preschool in Boston." *The Quarterly Journal of Economics*, 138(1): 363–411.

Hands, Angela, Sandy Gordon, Mohbubul Choudhury, Lola Toppin, and Jeremy Weingard. 2006. "Sure Start Children's Centres." National Audit Office.

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): 78–104.

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.

List, John, Dana Suskind, and Lauren Supplee. 2021. *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It.* Routledge.

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.

Løken, Katrine V., Kjell Erik Lommerud, and Katrine Holm Reiso. 2018. “Single mothers and their children: Evaluating a work-encouraging welfare reform.” *Journal of Public Economics*, 167: 1–20.

Melhuish, Edward, Jay Belsky, Alastair H Leyland, and Jacqueline Barnes. 2008*a*. “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.

Melhuish, Edward, Jay Belsky, Alastair H. Leyland, Angela Anning, David Hall, Jane Tunstill, Mog Ball, Pamela Meadows, Jacqueline Barnes, Martin Frost, Helena Romaniuk, Brian Bell, Marialena Trivella, Sapna Dave, and Mark Hibbett. 2005. “Early impacts of Sure Start Local Programmes on Children and Families.” National Evaluation of Sure Start.

Melhuish, Edward, Jay Belsky, Alastair H. Leyland, Angela Anning, David Hall, Jane Tunstill, Mog Ball, Pamela Meadows, Jacqueline Barnes, Martin Frost, Helena Romaniuk, Mark Hibbett, and Antero Malin. 2008*b*. “The impact of Sure Start Local Programmes on Three Year Olds and Their Families.” National Evaluation of Sure Start.

Melhuish, Edward, Jay Belsky, Alastair H. Leyland, Angela Anning, David Hall, Jane Tunstill, Mog Ball, Pamela Meadows, Jacqueline Barnes, Martin Frost, Mark Hibbett, and Andrew Cullis. 2010. “The impact of Sure Start Local Programmes on five year olds and their families.” National Evaluation of Sure Start, Department for Education.

Miller, Sarah, and Laura R. Wherry. 2019. “The long-term effects of early life Medicaid coverage.” *The Journal of Human Resources*, 54(3): 785–824.

Morpeth, Louise, Sarah Blower, Kate Tobin, Rod S. Taylor, Tracey Bywater, Rhiannon Tudor Edwards, Nick Axford, Minna Lehtonen, Carys Jones, and Vashti Berry. 2017. “The effectiveness of the Incredible Years pre-school parenting programme in the United Kingdom: a pragmatic randomised controlled trial.” *Child Care in Practice*, 23(2): 141–161.

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.

OECD. 2019. “OECD Family Database: PF1.6 Public spending by age of children.” Organisation for Economic Cooperation and Development.

Parenting, and Family Support Centre. 2025. “Triple-P evidence base.” The University of Queensland Australia.

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.

Roth, Jonathan. 2022. “Pretest with caution: Event-study estimates after testing for parallel trends.” *American Economic Review: Insights*, 4(3): 305–322.

Roth, Jonathan, Pedro H.C. Sant’Anna, Alyssa Bilinski, and John Poe. 2023. “What’s trending in difference-in-differences? A synthesis of the recent economics literature.” *Journal of Econometrics*, 235: 2218–2244.

Sammons, Pam, James Hall, Rebecca Smees, and Jenny Goff. 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) RR495.

Sandner, Malte, Stephan L Thomsen, and Libertad González. 2024. “Preventing Child Maltreatment: Beneficial Side Effects of Public Childcare.” *The Economic Journal*, 135(665): 321–353.

Schmidheiny, Kurt, and Sebastian Siegloch. 2023. “On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence and generalization.” *Journal of Applied Econometrics*, 38: 695–713.

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

Sylva, K., Goff J. Eisenstadt N. Smith T. Hall J. Evangelou M. Smith G., and P. Sammons. 2015. “Organisation, services and reach of children’s centres: Evaluation of children’s centres in England (ECCE, Strand 3).” Department for Education Report no. RB433.

Thompson, Owen. 2017. “Head Start’s Long-Run Impact: Evidence from the Program’s Introduction.” *Journal of Human Resources*, 53(4): 1100–1139.

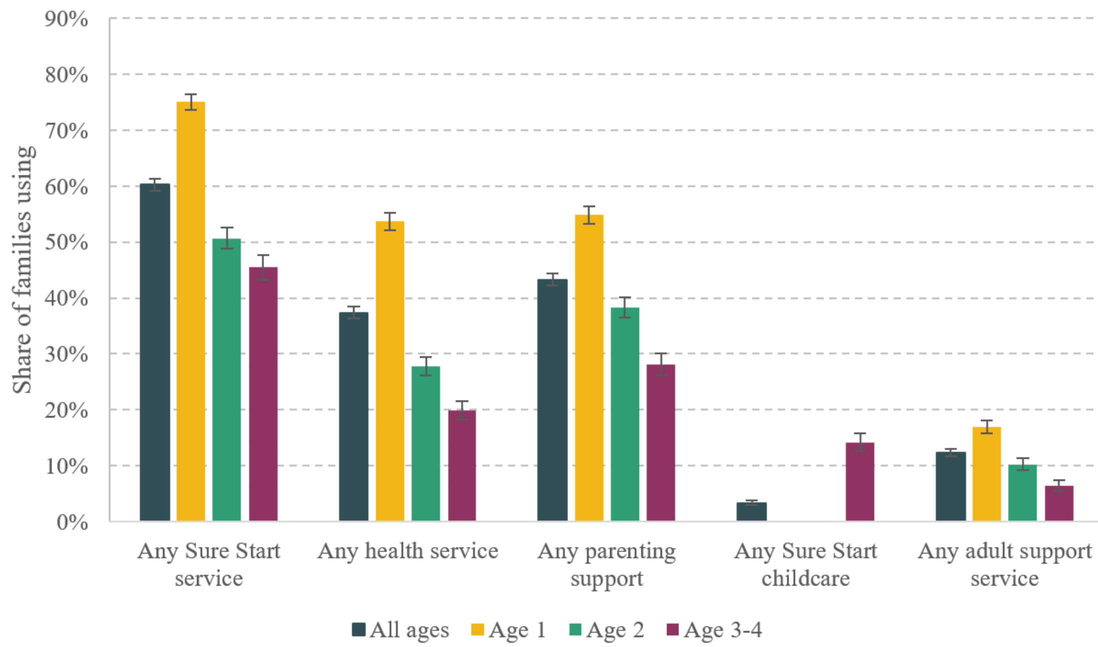
van den Berg, Gerard J., and Bettina M. Siflinger. 2022. “The effects of a daycare reform on health in childhood – Evidence from Sweden.” *Journal of Health Economics*, 81: 102577.

Walters, Christopher R. 2015. “Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start.” *American Economic Journal: Applied Economics*, 7(4): 76–102.

Welshman, John. 2010. “From Head Start to Sure Start: Reflections on Policy Transfer.” *Children and Society*, 24: 89–99.

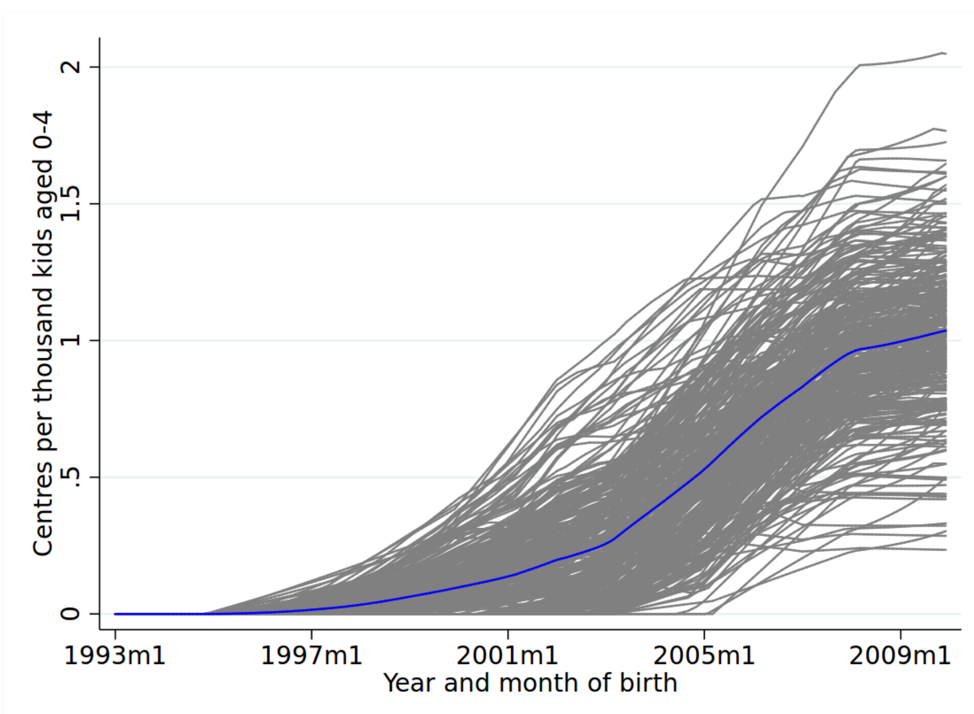
Wherry, Laura R., Sarah Miller, Robert Kaestner, and Bruce D. Meyer. 2018. “Childhood Medicaid coverage and later-life health care utilization.” *The Review of Economics and Statistics*, 100: 287–302.

Figure 1: Take-up of Sure Start services in the early 2010s (Phase 1 and 2 centers)



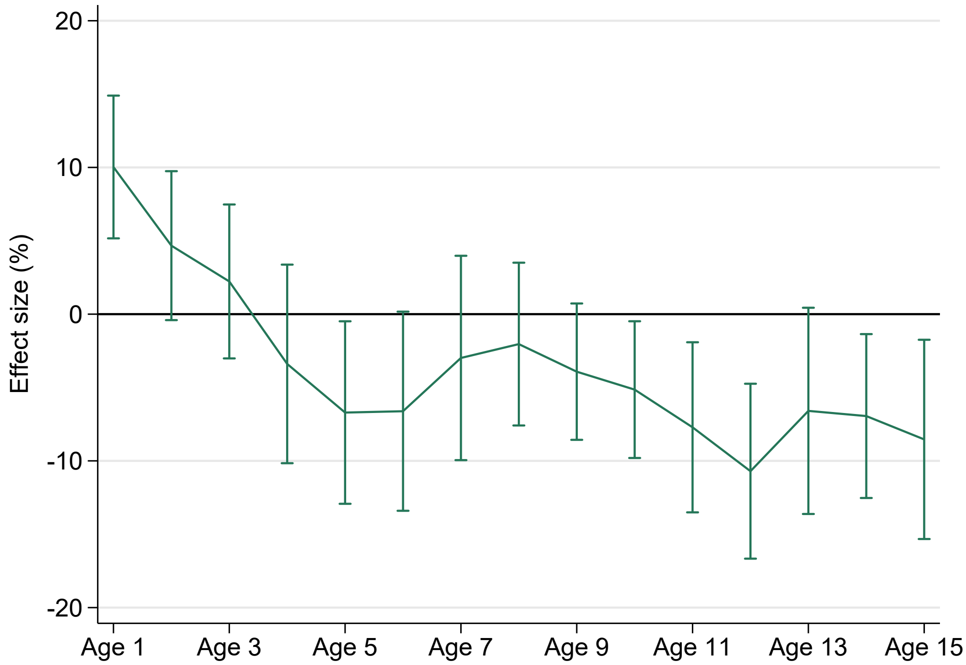
Notes: The figure shows the probability that families in the ECCE sample report making use of Sure Start, overall and for each of the four main types of services, by the age of the focus child. This is based on pooling all three waves of ECCE data. Data source: Goff and Chu (2013).

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



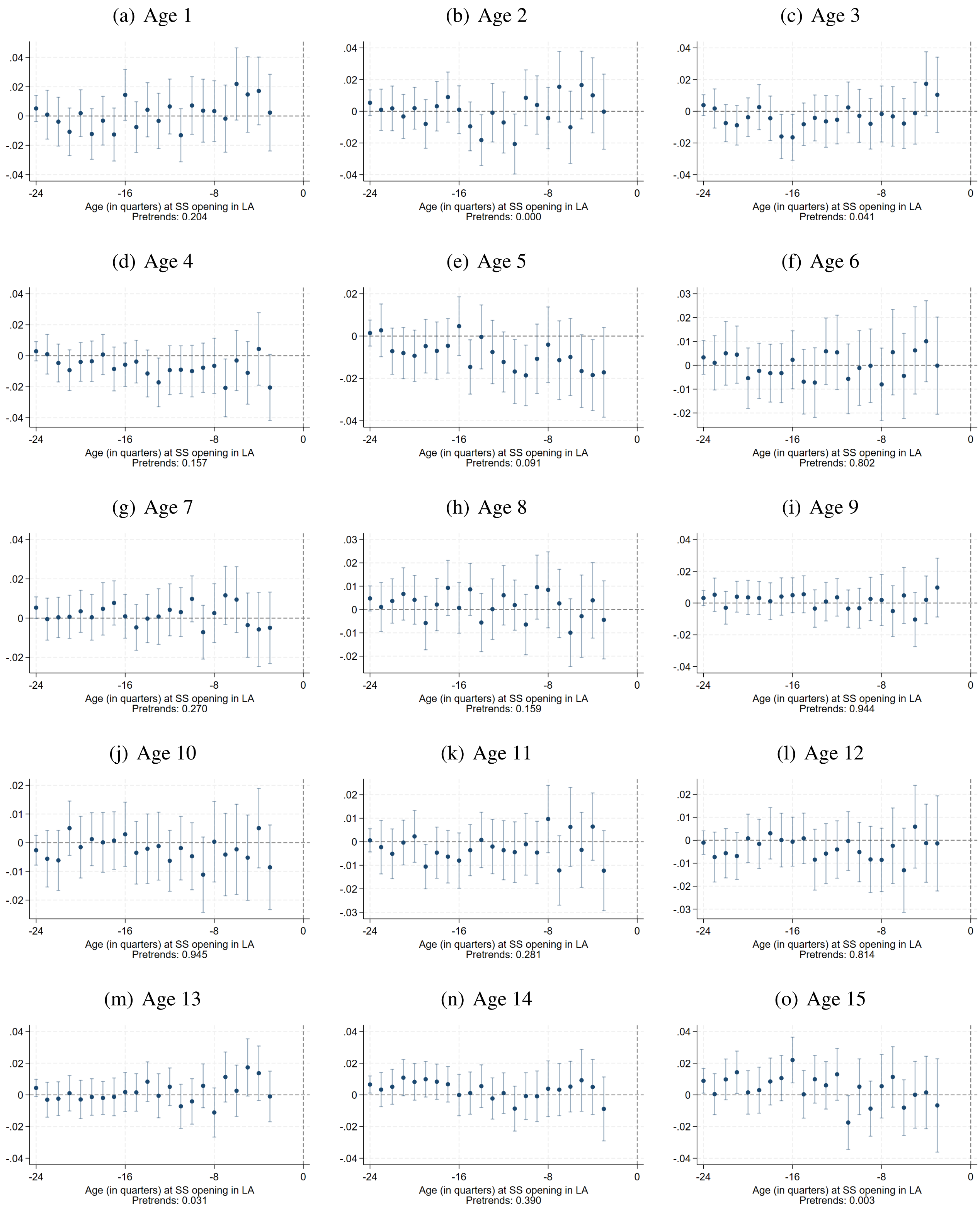
Notes: Each gray line represents one of 323 Local Authority (LA) districts in England (excluding the Isles of Scilly, City of London, and West Somerset). The blue line shows the average for all of England. The lines plot the average Sure Start coverage (centers per thousand children aged 0-4 in the LA) 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.

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



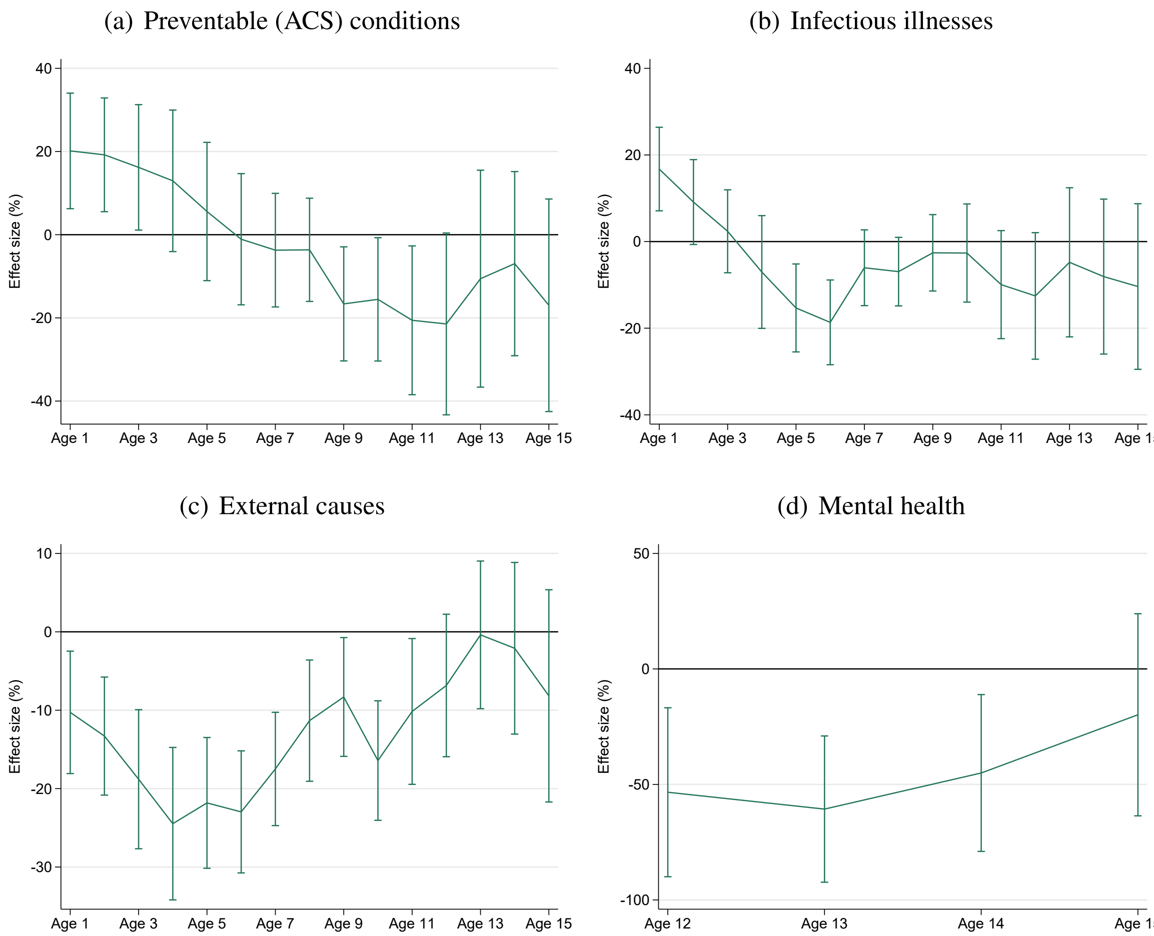
Note: Effect sizes are constructed by rescaling the estimates by the pre-Sure Start baseline probability of a hospitalization at each age. Vertical bars indicate 90% confidence intervals. Source: Authors' calculations using data from the Hospital Episode Statistics inpatient data (1997-2017) and the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

Figure 4: Regression of hospitalizations on quarterly leads of Sure Start treatment, in untreated cohorts



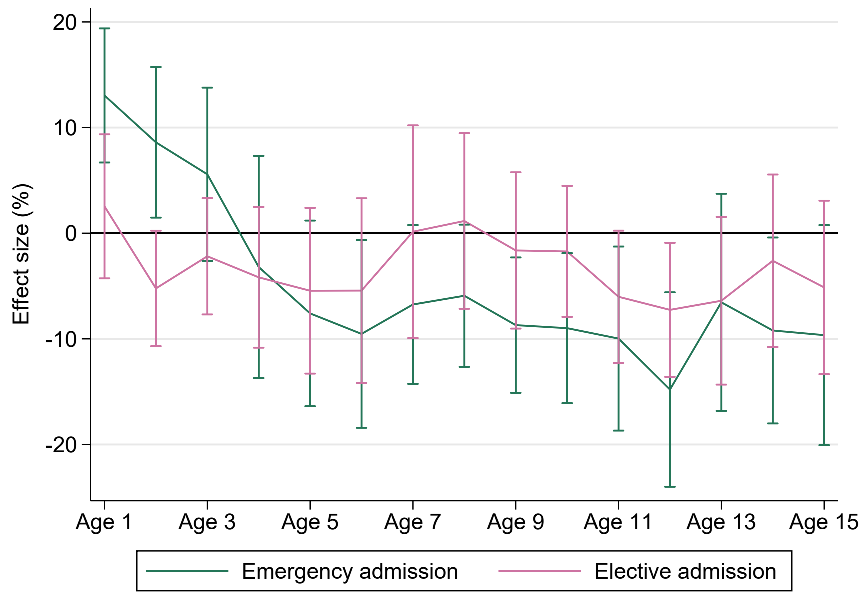
Note: Each panel regresses hospitalizations at one age on quarterly leads of Sure Start treatment. To avoid collinearity between leads and lags induced by using an averaged treatment measure, we focus on treatment experienced in the month before the child's fifth birthday (or the month before the relevant birthday for outcomes at earlier ages). We follow Schmidheiny and Siegloch (2023) in binning endpoints and Borusyak, Jaravel and Spiess (2024) in dropping one lead to avoid underidentification. Each figure presents the p-value of an F text for the joint significance of the coefficients. 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 5: Effect of an increase in Sure Start coverage on probability of hospitalization for specific causes, re-scaled by baseline probability



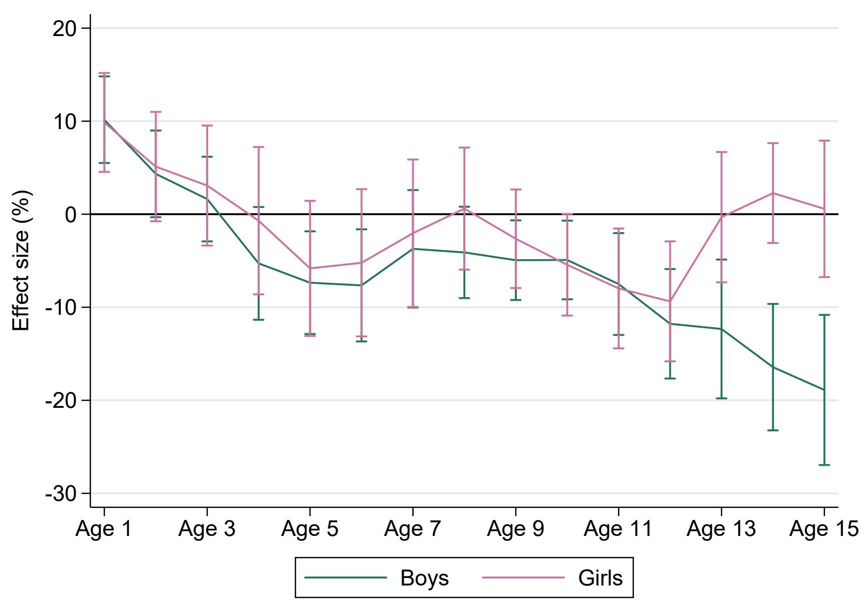
Note: Figure shows coefficients from separate regressions for each outcome age. Coefficients are re-scaled by the baseline (1996) mean for each age. Vertical bars indicate 90% confidence intervals. Cause-specific results are based on the primary diagnosis at the time of admission. See Blunt (2013) for 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.

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



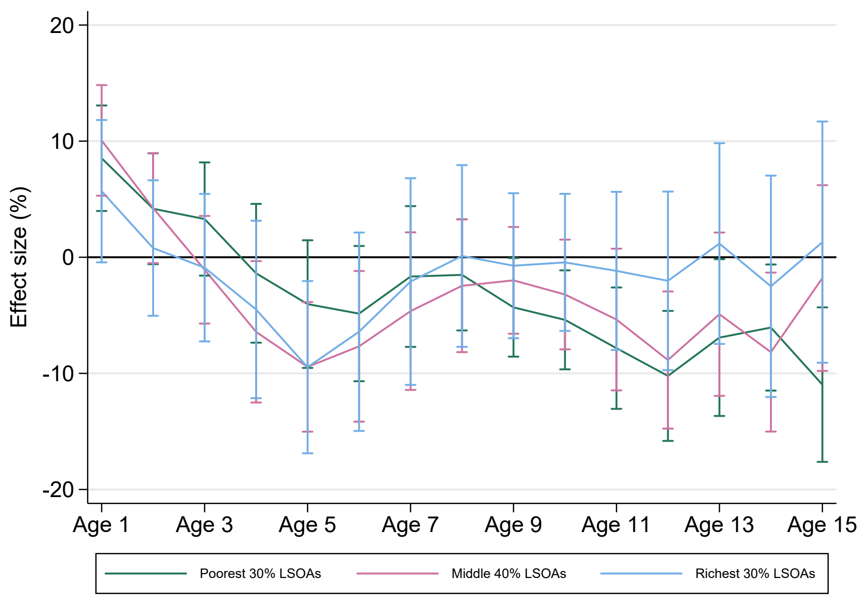
Note: The figure shows coefficients from separate regressions for each outcome age. Coefficients are re-scaled by the baseline (1996) mean for each age. Vertical bars indicate 90% confidence intervals.

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.

Figure 8: 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.

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*** (0.0075)+++	0.0095 (0.0063)	0.0040 (0.0057)	-0.0057 (0.0069)	-0.0109* (0.0061)	-0.0095 (0.0059)	-0.0038 (0.0053)	-0.0024 (0.0039)	-0.0044 (0.0032)	-0.0055* (0.0031)	-0.0084** (0.0038)+	-0.0125*** (0.0042)+++	-0.0081 (0.0052)	-0.0091** (0.0045)	-0.0120** (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 Latest cohort	Apr.96 Dec.06	Apr.95 Dec.06	Apr.94 Dec.06	Apr.93 Dec.06	Jan.93 Dec.06	Jan.93 Dec.06	Jan.93 Dec.06	Jan.93 Dec.06	Jan.93 Dec.06	Jan.93 Dec.06	Jan.93 Mar.06	Jan.93 Mar.05	Jan.93 Mar.04	Jan.93 Mar.03	Jan.93 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).

Table 2: 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 3: Potential effects of Sure Start on children's hospitalizations

Service	Mechanism	Potential impact: Early years (c. age 1-4)	Potential impact: Childhood (c. age 5-10)	Potential impact: Adolescence (c. age 11-15)
Health services & information about health and safety	Screening and referrals to healthcare	? Preventable	↓ Preventable	↓ Preventable
	Safer home environments	↓ External (poisoning, injuries)	⊘	⊘
Parenting support and parent-child activities	Better emotional and behavioural development	↓ External (accidents, injuries)	↓ External (accidents, injuries)	↓ External (accidents, injuries)
		↓ Mental health *(Not measured)*	↓ Mental health *(Not measured)*	↓ Mental health
Childcare and group sessions	Stronger immune system	↑ Infectious illness	↓ Infectious illness	⊘
Adult support	Parents move into paid work: Higher family income	↓ All causes	↓ All causes	↓ All causes
	Parents move into paid work: Less time at home	↑ All causes	↑ All causes	↑ All causes

Table 4: 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.

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

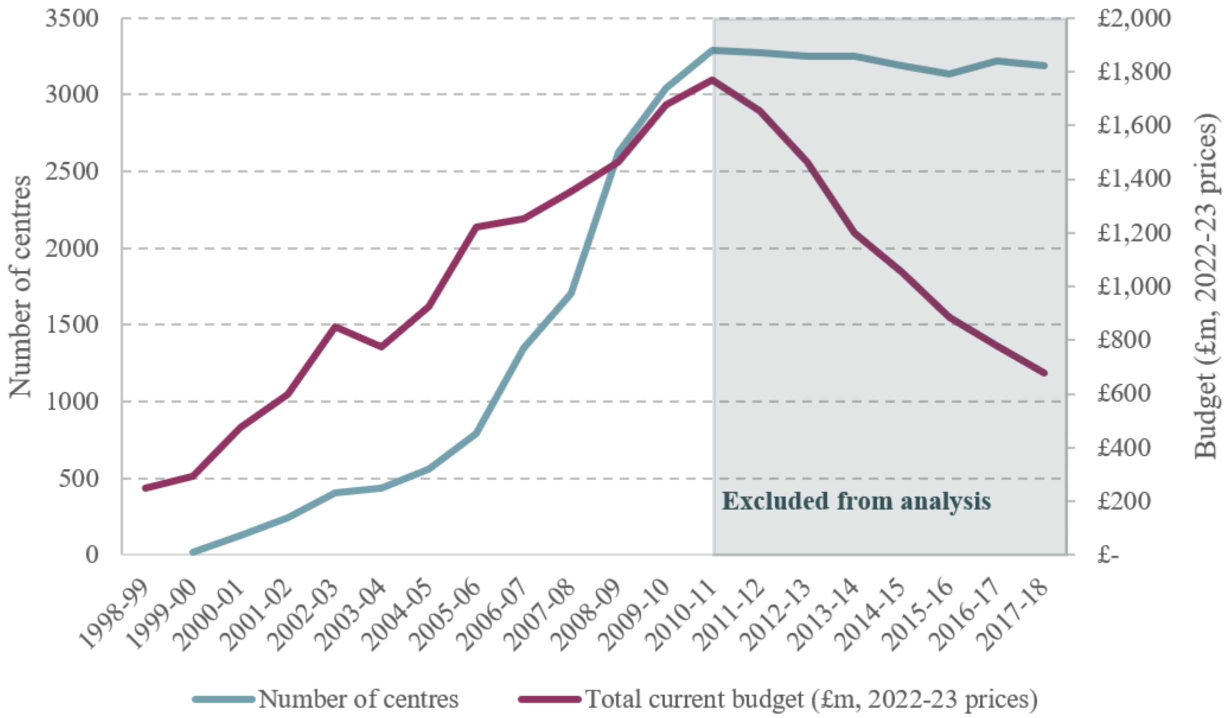
Sarah Cattan, Gabriella Conti, Christine Farquharson, Rita Ginja and Maud Pecher

May 19, 2025

MATERIAL FOR ONLINE APPENDIX

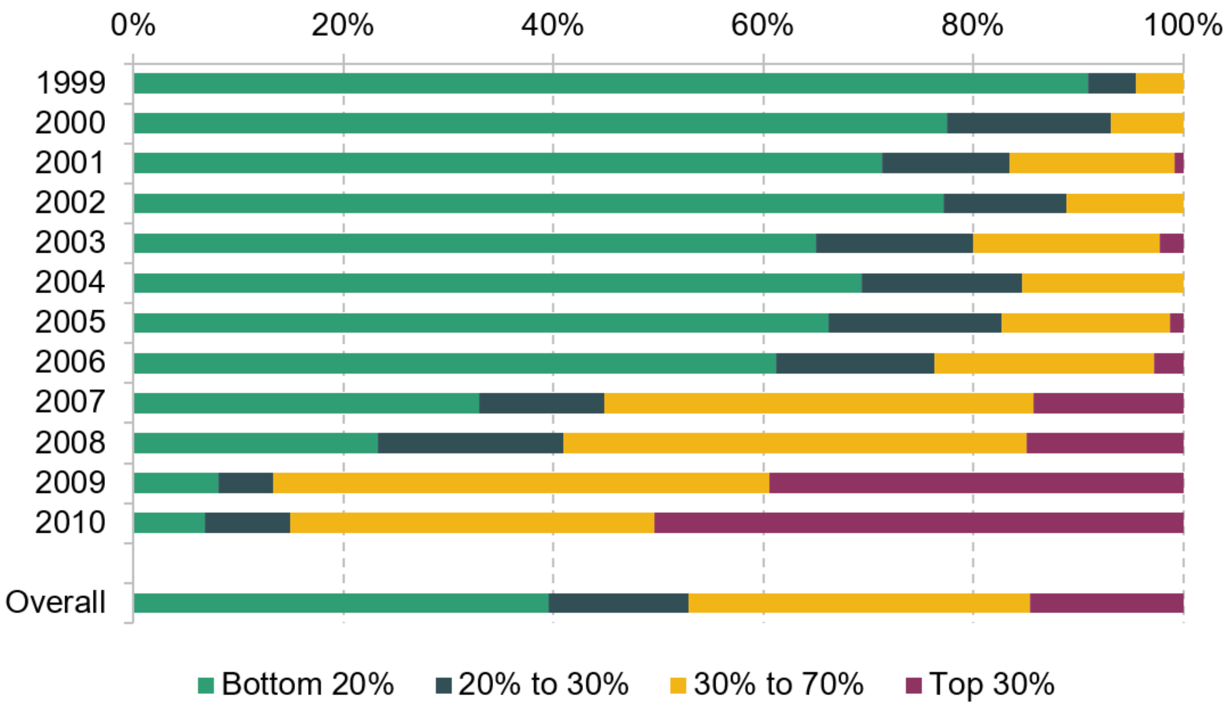
A Appendix Tables and Figures

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



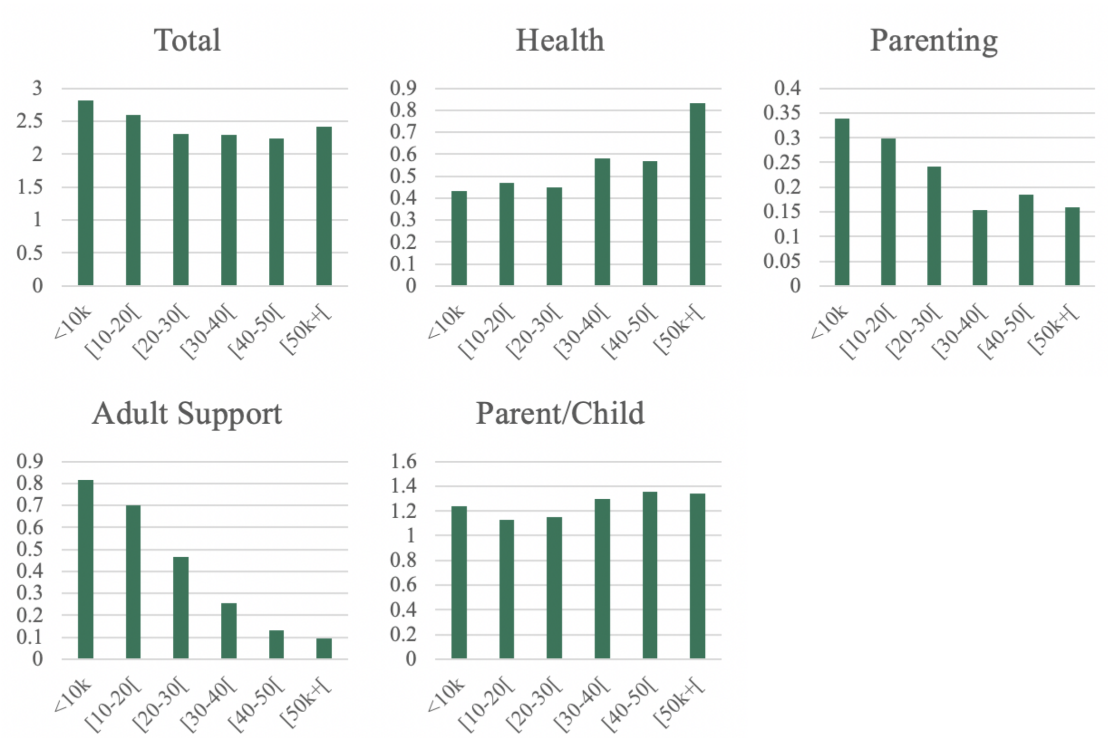
Note: The number of centers is based on centers observed in data received from the Department for Education. Since the treatment of arrangements like satellite sites was not always consistent, these numbers might not exactly match other data sources. We assume that a Sure Start Children's center (SSCC) 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. Data on centers from 2010 onward incorporates official center closures but not consolidations, service reductions or other types of 'hollowing out'. Budget data refers to the current (not capital) budget. Source: Authors' calculations using data provided by the Department for Education. Sure Start budget data from (Britton, Farquharson and Sibieta, 2019), up-rated using the June 2023 GDP deflator for the UK.

Figure A.2: Share of Sure Start centres opened each year, by deprivation



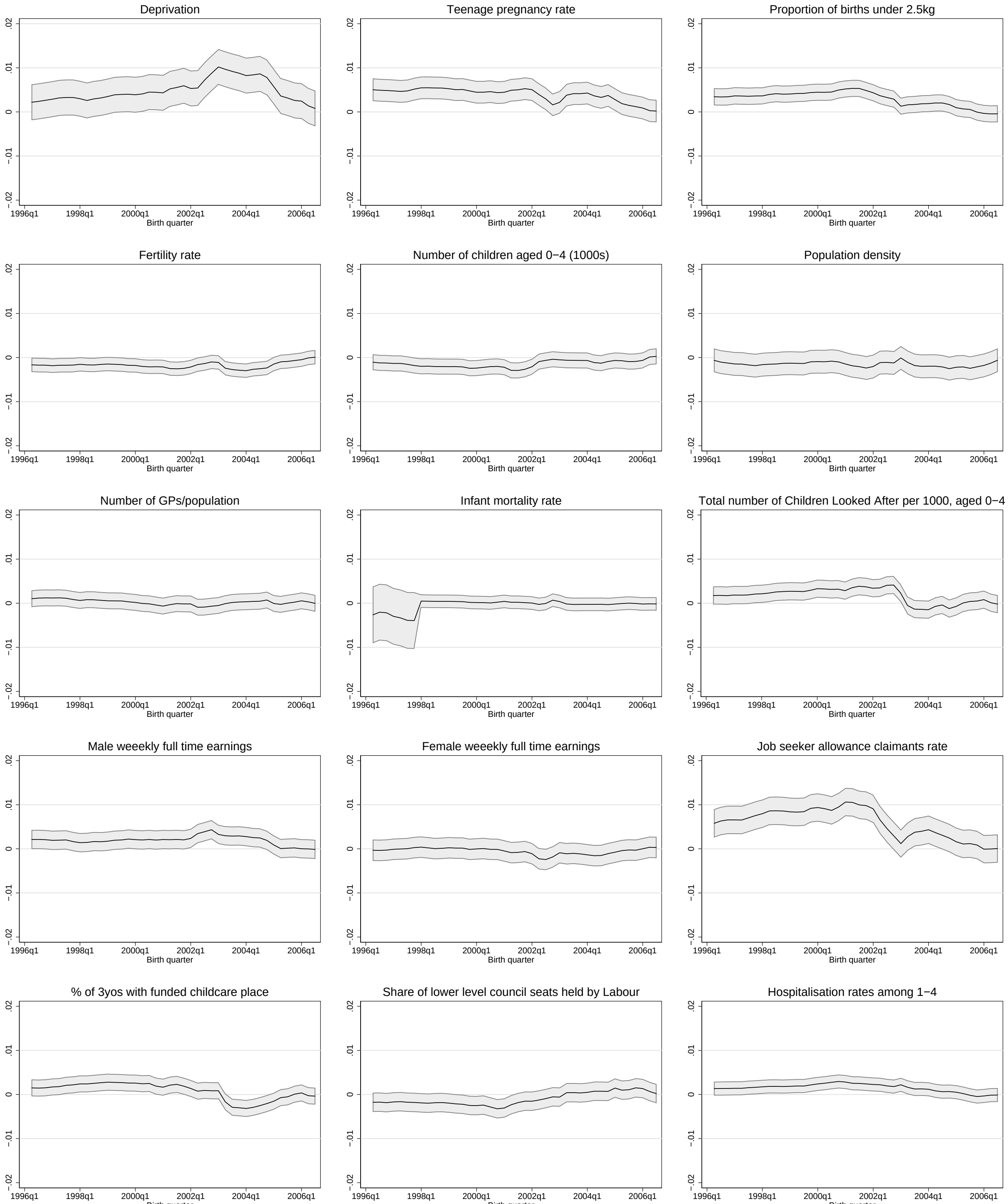
Note: 'Bottom 20%' refers to neighbourhoods (lower layer super output areas, LSOAs) in the bottom 20% of the national 2004 Index of Multiple Deprivation ranking, i.e. the most disadvantaged. Other categories are similarly defined, with 'Top 30%' incorporating the least disadvantaged areas.

Figure A.3: Hours spent per week at different Sure Start services by family income, 2011



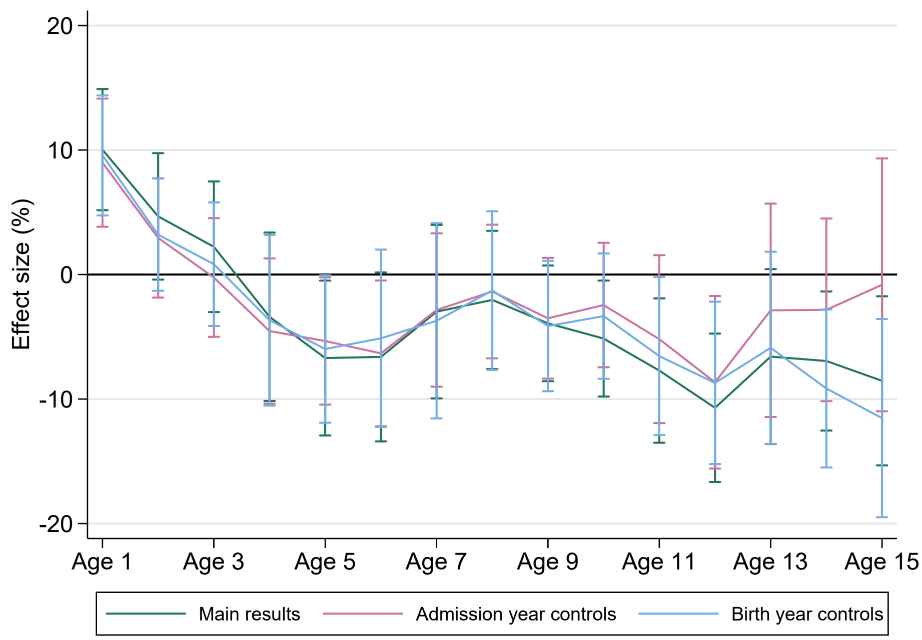
Note: The figure is based on information collected by the Evaluation of Children's centers in England (ECCE) in 2011 on hours spent per week by families using different services.

Figure A.4: Regression of the change in Sure Start coverage on baseline Local Authority characteristics, 1998 - 2006



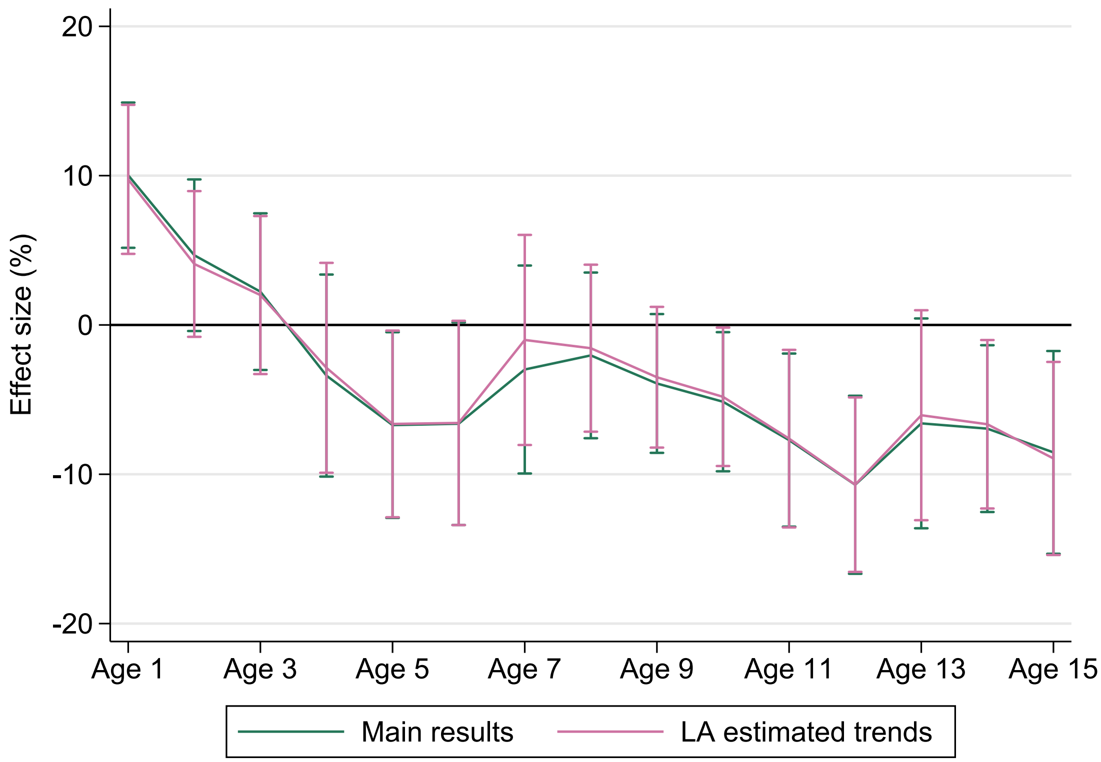
Note: These figures plot the coefficients obtained from a regression of the changes in Sure Start coverage on Local Authority specific baseline characteristics (measured in 1998) interacted with quarter-year dummies, controlling for Local Authority fixed effects. Every characteristic has been standardized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

Figure A.5: Effect of an increase in Sure Start coverage on probability of any hospitalization, rescaled by baseline probability: Robustness to inclusion of time-varying controls



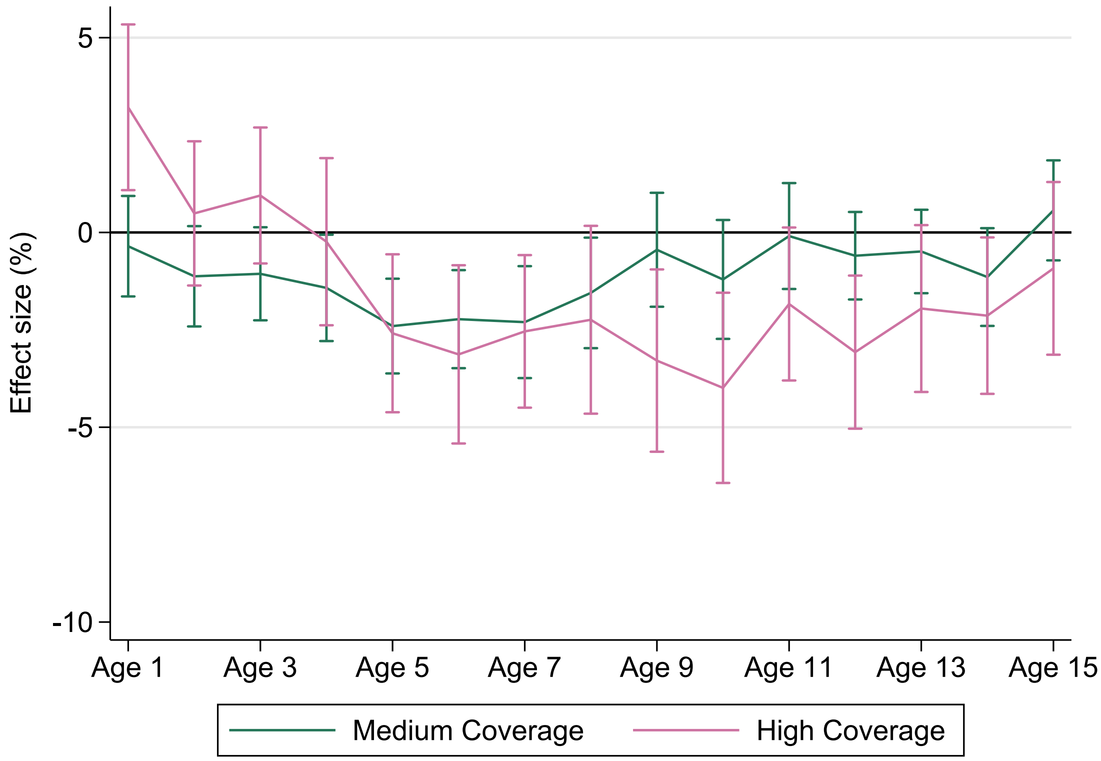
Note: Figure shows coefficients from separate regressions for each outcome age. Coefficients are rescaled by the baseline (1996) mean for each age. Specification including time-varying controls contains controls for: the teenage conception rate; the share of births with low birth weight; the total period 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 C.1.

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



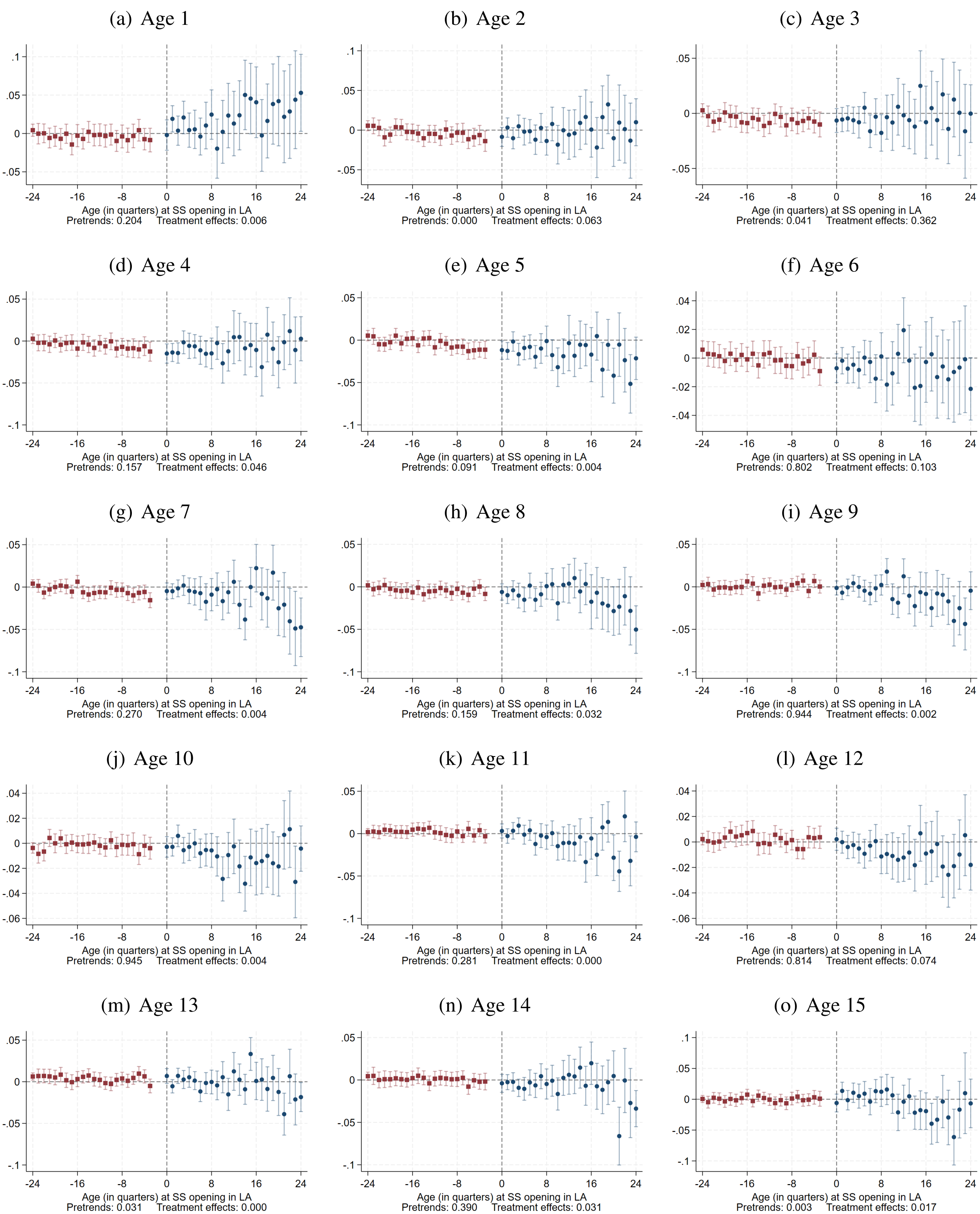
Note: Effect sizes are constructed by re-scaling the estimates by the pre-Sure Start (1996) baseline probability of a hospitalization at each age. Vertical bars indicate 90% confidence intervals. Results with LAD estimated trends additionally control for a local authority-specific linear time trend, 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: Non-linear estimates



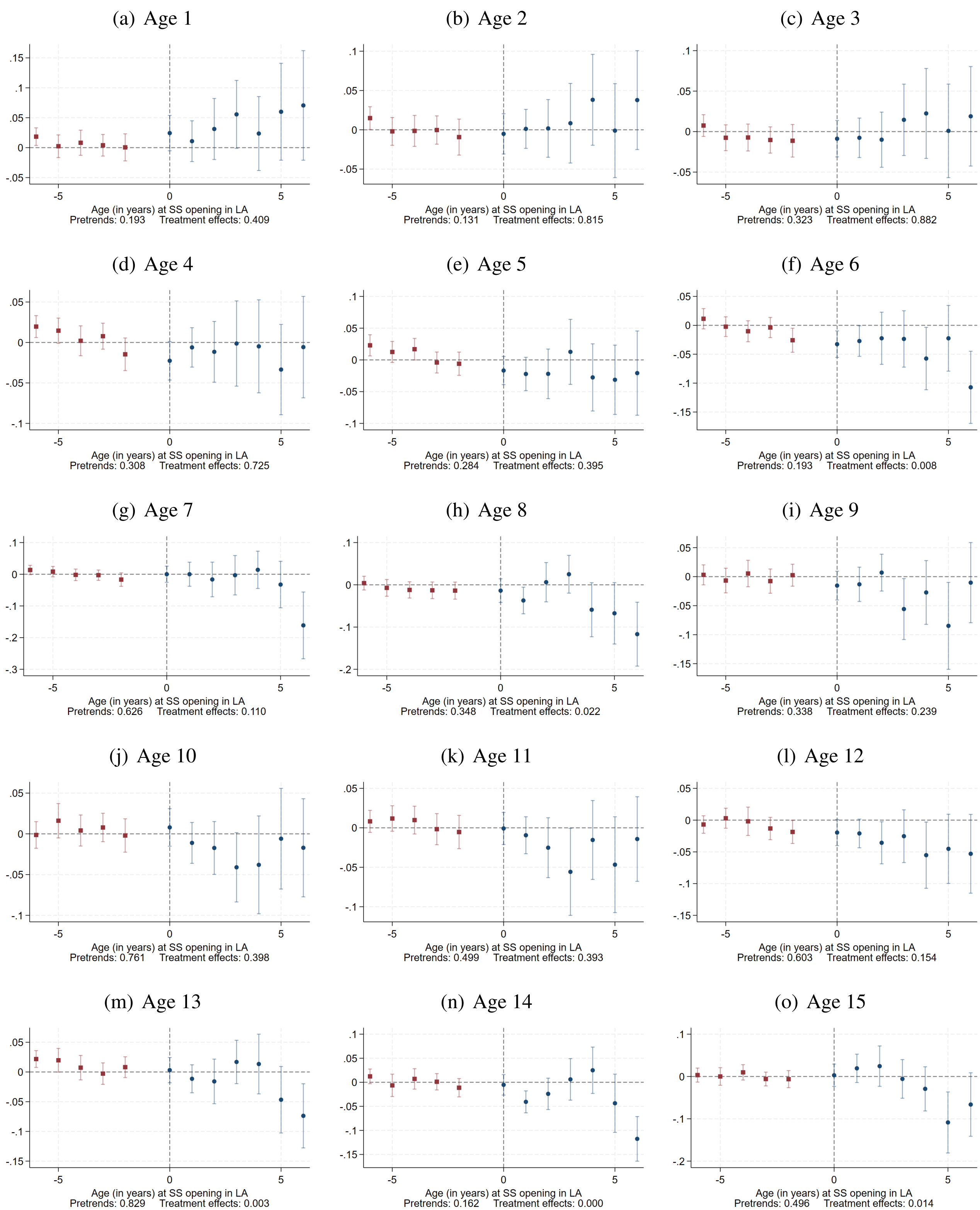
Note: The figure shows coefficients from separate regressions for each outcome age. Treatment is a pair of indicators for whether the cell experienced medium treatment (strictly positive coverage, but less than 0.25 centers per thousand children) or high treatment (more than 0.25 centers per thousand 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.8: Regression of hospitalizations on quarterly leads and lags of Sure Start coverage



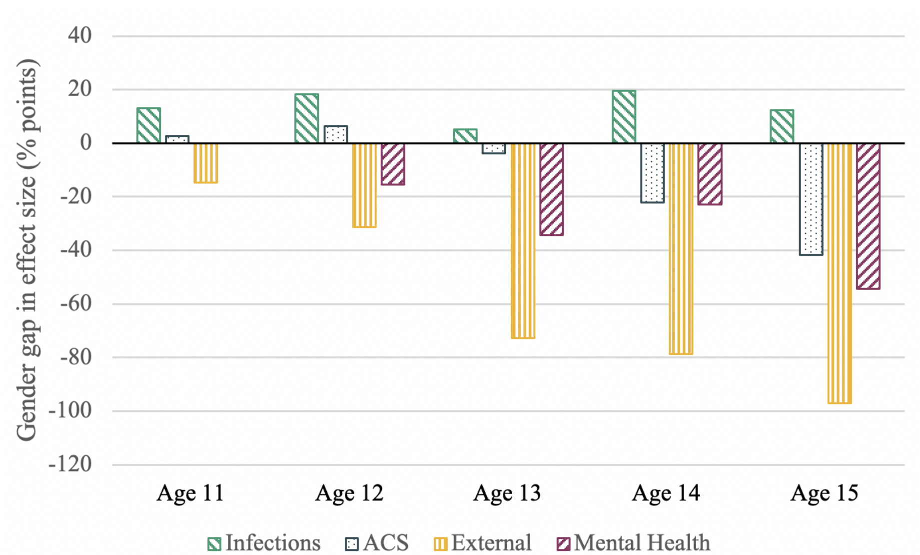
Note: Each panel regresses hospitalizations at one age on quarterly leads and lags of Sure Start treatment. To avoid collinearity between leads and lags induced by using an averaged treatment measure, we focus on treatment experienced in the month before the child's fifth birthday (or the month before the relevant birthday for outcomes at earlier ages). We bin endpoints and drop one lead to avoid under-identification. Each figure presents the p-value of an F text for the joint significance of the treatment coefficients.We test the significance of our pre-trends in a separate specification just using untreated cohorts (Figure 4); for ease, we repeat these p-values in this Figure as well.

Figure A.9: Regression of hospitalizations on yearly leads and lags of Sure Start coverage



Note: Each panel regresses hospitalizations at one age on yearly leads and lags of Sure Start treatment. To avoid collinearity between leads and lags induced by using an averaged treatment measure, we focus on treatment experienced in the month before the child's fifth birthday (or the month before the relevant birthday for outcomes at earlier ages). We bin endpoints and drop one lead to avoid under-identification. Each figure presents the p-value of an F text for the joint significance of the treatment coefficients. We test the significance of our pre-trends in a separate specification just using untreated cohorts; this p-value is reported in each figure as well.

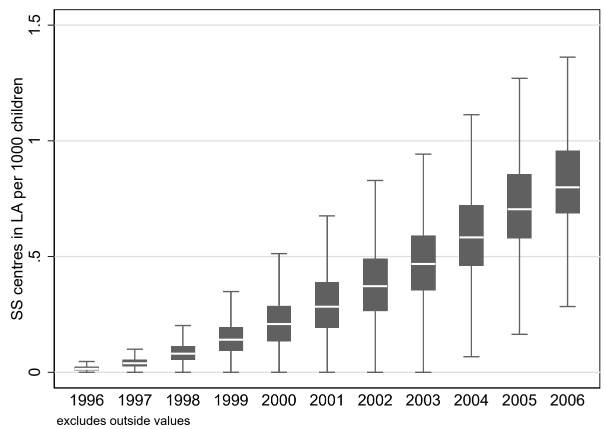
Figure A.10: Gender gap in the Effect of an increase in Sure Start coverage on probability of cause-specific hospitalizations



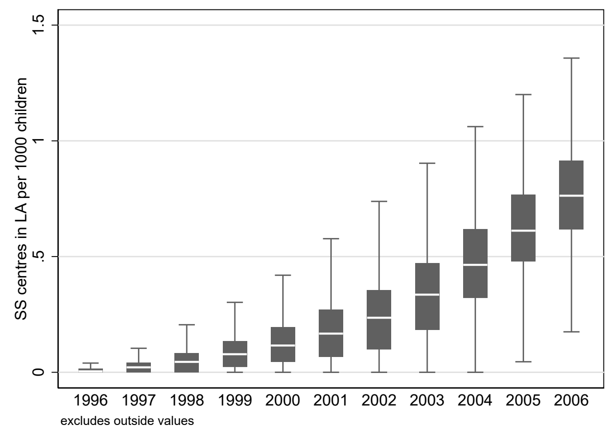
Note: The Figure shows the percentage point difference the estimated effect size of Sure Start on the probability of hospitalization between boys and girls. The difference in effect size between both genders is statistically significant at the 90% level at ages 11-12 for infections, age 15 for ACS and ages 11-15 for external. Table A.14 and Table A.15 display the original cause-specific point estimates and p-values by gender for ages 1-15.

Figure A.11: Average coverage over the first 60 months of life, by local authority and month and year of birth: By level of deprivation

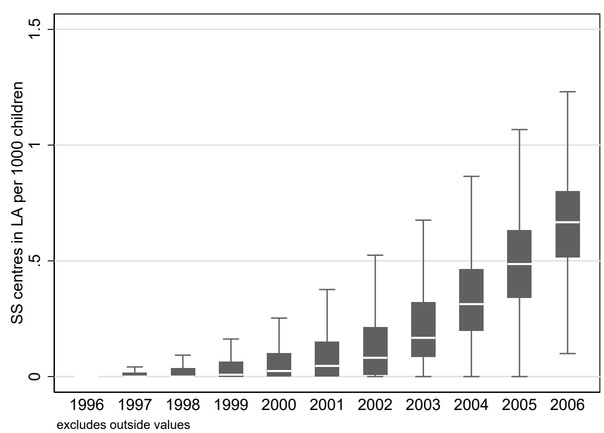
(a) Poorest 30% LSOAs



(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.

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)
N	8882	8880	8882	8880
Controls and region FE?		Y		Y

Source: Millennium Cohort Study (MCS), Wave 1 (9 months old). Data citation: University of London. Institute of Education. Centre for Longitudinal Studies. (2017). Millennium Cohort Study: First Survey, 2001-2003. [data collection]. 12th Edition. UK Data Service. SN: 4683, DOI: 10.5255/UKDA-SN-4683-4.

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. *, ** 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) Parenting support	(4) Childcare	(5) Adult support
Child age (reference: age 1)					
Age 2	-0.233***	-0.261***	-0.152***	-0.000	-0.066***
	(0.010)	(0.010)	(0.010)	(0.004)	(0.007)
Age 3-4	-0.281***	-0.342***	-0.251***	0.141***	-0.112***
	(0.011)	(0.011)	(0.012)	(0.004)	(0.008)
Family SES index, interacted with:					
Child age 1	0.022***	0.016***	0.039***	0.001	-0.014***
	(0.006)	(0.006)	(0.006)	(0.002)	(0.004)
Child age 2	-0.054***	-0.050***	-0.043***	0.001	-0.040***
	(0.009)	(0.009)	(0.009)	(0.003)	(0.007)
Child age 3-4	-0.047***	-0.025**	-0.028***	-0.034***	-0.012
	(0.010)	(0.010)	(0.011)	(0.004)	(0.007)
Child is female	-0.013	-0.005	-0.020**	-0.004	-0.008
	(0.009)	(0.008)	(0.009)	(0.003)	(0.006)
Child ethnicity (reference: white)					
Asian	-0.066***	-0.106***	-0.075***	-0.005	0.025**
	(0.015)	(0.014)	(0.015)	(0.005)	(0.010)
Black	0.051***	0.001	0.076***	0.013**	0.085***
	(0.018)	(0.018)	(0.019)	(0.007)	(0.013)
Mixed	-0.002	-0.097***	0.056***	0.005	-0.007
	(0.018)	(0.017)	(0.018)	(0.006)	(0.013)
Other	-0.040	-0.090***	-0.036	-0.002	-0.003
	(0.031)	(0.031)	(0.032)	(0.011)	(0.022)
Mother has partner	0.024**	0.021*	0.032**	-0.010**	-0.045***
	(0.012)	(0.012)	(0.013)	(0.004)	(0.009)
Mother age at wave 1 (reference: <25)					
25-34	0.001	0.009	0.002	0.000	-0.027***
	(0.012)	(0.012)	(0.013)	(0.004)	(0.009)
35 and older	-0.020	-0.032**	-0.002	0.005	-0.030***
	(0.014)	(0.014)	(0.015)	(0.005)	(0.010)
Observations	11,835	11,835	11,835	11,835	11,835
R-squared	0.078	0.106	0.053	0.116	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: Effect of an increase in Sure Start coverage on probability of hospitalization for any cause: Robustness to controlling for and weighting by population

	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 14	Age 15
Panel A: Main results															
SS coverage	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)
Panel B: No population controls															
SS coverage	0.0232***	0.0070	0.0004	-0.0107	-0.0179***	-0.0170***	-0.0113*	-0.0097**	-0.0118***	-0.0125***	-0.0143***	-0.0174***	-0.0120*	-0.0120*	-0.0154**
	(0.0074)	(0.0063)	(0.0058)	(0.0073)	(0.0065)	(0.0065)	(0.0060)	(0.0046)	(0.0040)	(0.0039)	(0.0047)	(0.0051)	(0.0061)	(0.0053)	(0.0065)
Panel C: Weighting by population															
SS coverage	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)
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. Panel A reproduces our main results from Table 1. In Panel B, results are presented without time-varying controls for LSOA population. Panel C weights the data by the number of children in each LSOA. '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.5: 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 and local unemployment															
SS Cov	0.0246***	0.0071	0.0027	-0.0072	-0.0120**	-0.0105*	-0.0047	-0.0034	-0.0051	-0.0061**	-0.0084**	-0.0128***	-0.0091*	-0.0100**	-0.0126**
	(0.0074)	(0.0061)	(0.0056)	(0.0068)	(0.0060)	(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 and local unemployment															
SS Cov	0.0244***	0.0069	0.0027	-0.0072	-0.0120**	-0.0105*	-0.0047	-0.0035	-0.0051	-0.0062**	-0.0085**	-0.0130***	-0.0094*	-0.0105**	-0.0132**
	(0.0073)	(0.0061)	(0.0056)	(0.0068)	(0.0060)	(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
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 and local unemployment conditions 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. Local unemployment is measured by the Job Seeker Allowance Claimant Rate. 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.6: 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.7: 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.8: Comparison of binary treatment effect estimates using the TWFE and Borusyak, Jaravel and Spiess (2024) 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 (2024) 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.9: 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.10: 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*** (0.0011)+++	-0.0025** (0.0012)++	-0.0024** (0.0010)++ 2.4	-0.0008 (0.0007)	-0.0002 (0.0006)	-0.0001 (0.0004)	-0.0004 (0.0004)	0.0003 (0.0003)	0.0006* (0.0003)	-0.0003 (0.0003)	0.0003 (0.0003)	0.0006 (0.0005)	0.0004 (0.0009)	0.0015 (0.0014)	-0.0009 (0.0019)
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 (0.0012)	-0.0030** (0.0012)++	-0.0033*** (0.0010)+++	-0.0056*** (0.0013)+++	-0.0054*** (0.0011)+++	-0.0055*** (0.0010)+++	-0.0035*** (0.0009)+++	-0.0026*** (0.0009)+++	-0.0022*** (0.0009)+++	-0.0033*** (0.0010)+++	-0.0026** (0.0013)	-0.0024* (0.0014)	-0.0006 (0.0014)	-0.0018 (0.0018)	-0.0027 (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* (0.0004)	-0.0009** (0.0004)	-0.0007* (0.0004)	-0.0008 (0.0006)	-0.0016*** (0.0005)+++	-0.0028*** (0.0006)+++	-0.0009* (0.0005)	-0.0012** (0.0005)+	-0.0013*** (0.0005)+++	-0.0013*** (0.0005)+++	-0.0018*** (0.0007)+++	-0.0010 (0.0007)	0.0005 (0.0009)	0.0008 (0.0010)	0.0005 (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 (0.0008)	-0.0016* (0.0008)	-0.0010 (0.0006)	-0.0028*** (0.0007)+++	-0.0019*** (0.0006)+++	-0.0005 (0.0004)	-0.0006* (0.0004)	-0.0007 (0.0004)	-0.0006** (0.0003)	-0.0006* (0.0003)	0.0000 (0.0004)	-0.0005 (0.0005)	-0.0004 (0.0005)	-0.0011 (0.0007)	-0.0017* (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.11: 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.12: 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.13: 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).

Table A.14: 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.15: 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.

B 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 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.[1] 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 B.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.

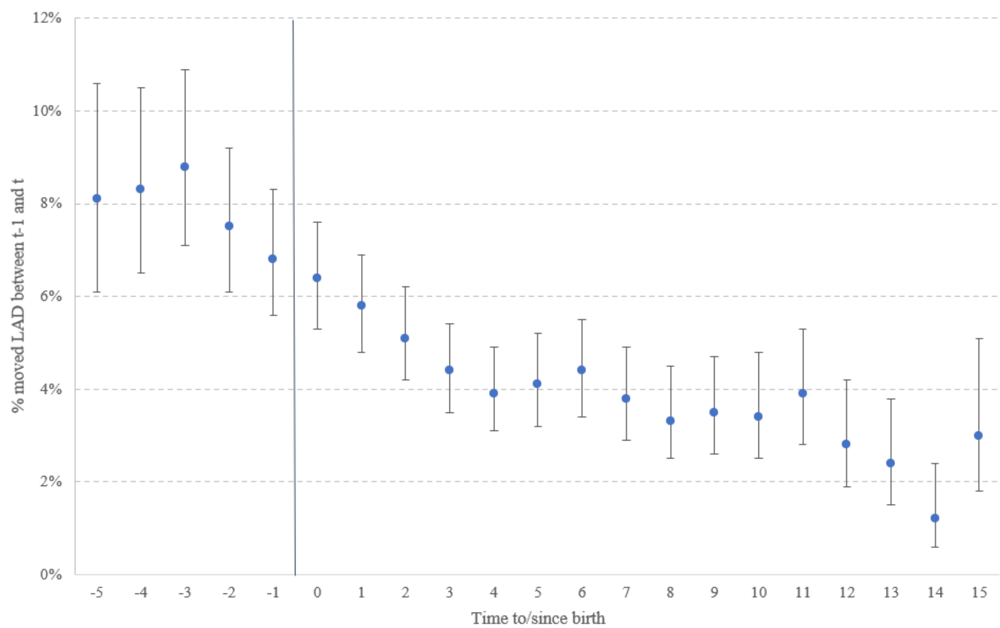
We also find that inter-LA mobility does not systematically relate to Sure Start availability. Table B.1 shows that children living at time t in LAs with greater access to Sure Start are no more

[1]Data citation: University of Essex, Institute for Social and Economic Research. (2020). Understanding Society: Waves 1-10, 2009-2019 and Harmonised BHPS: Waves 1-18, 1991-2009: Secure Access. [data collection]. 11th Edition. UK Data Service. SN: 6676, http://doi.org/10.5255/UKDA-SN-6676-11.

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.

Finally, in Table B.2, we analyse the relationship between Sure Start coverage and inter-LA mobility after the child has aged out of Sure Start treatment. This speaks to the extent to which there is *differential* measurement error in our main hospitalization-based results. We find that families with children age 6-7 are slightly less likely to have moved LA if they had experienced higher Sure Start coverage in the early years, but these impacts are generally small.

Figure B.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 B.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.

Table B.2: Effect of Sure Start coverage on likelihood of moving LAD

	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 14	Age 15
SS Cov.	-0.040**	-0.041*	-0.029	-0.006	0.012	0.030	0.044	-0.004	-0.058	0.121
	(0.018)	(0.024)	(0.031)	(0.045)	(0.051)	(0.078)	(0.089)	(0.120)	(0.153)	(0.229)
Observations	1,005	958	894	827	806	744	666	607	530	453
R-squared	0.003	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001

Note: Each column regresses an indicator for whether a parent had moved Local Authority since their child was five years old on the Sure Start coverage experienced by that child over their first five years (measured based on the LA at age 5). Standard errors are clustered at the LA level (based on the LA at age 5). *, ** and *** indicate significance at the 10%, 5% and 1% level respectively. Source: Authors' calculations using data from the British Household Panel Survey, 1991-2009.

C Sure Start rollout

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 subsection 4.2.

Table C.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 C.1: Covariates used in the rollout analysis

Category	Variable	Source	Years	Geography	Baseline mean [1998]
Deprivation	Percentile of rank distribution of Index of Local Deprivation	Department of Environment, Trade, and the Regions[1]	1998	LAD	50.0
Health Indicators	Under-18 conception rate (conceptions/1,000 women aged 15-17)	Child and Maternal Health Intelligence Network[2]	1998-2018	LAD	4.35
	Proportion of births below 2.5kg	ONS Vital Statistics[3]	1991-2018 (interpolated in 2008 and 2009)	LAD	0.071
Potential Demand	Total period fertility rate	ONS Vital Statistics[4]	1990-2018	LAD	1.763
for services	Density	ONS Population Density[5]	1990-2018	LAD	1.487
	% of primary school pupils with English as an Additional Language	National Association for Language Development in the Curriculum (NALDIC)	1999; 2004-2018	County	0.061 [1999]
	Children Looked After per thousand (under 1)	Department for Education	1992-2018	County	2.70
	Children Looked After per thousand (1 to 4)	Department for Education	1992-2018	County	3.12
Labour Market	Rate of Jobseekers Allowance receipt	Jobseekers Allowance[7]	1992-2018	LAD	0.030
Pre-Existing Services	Number of GPs per 1,000 population	Constructed with HSCIC data[10]	1990-2018	LAD	0.213
	Number of JobcentrePlus per 1,000 population	Department for Work and Pensions	2001-2018	LAD	0.001 [2001]
	Free entitlement take-up rate among 3 and 4-year-olds	Department for Education Statistical Returns	1997-2018	County	0.613

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.

D 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.[2] 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. (2022)). 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.[3]

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.

[2]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.

[3]Data citation: Office for National Statistics, Social Survey Division, Northern Ireland Statistics and Research Agency, Central Survey Unit. (2021). Quarterly Labour Force Survey, 1992-2020: Secure Access. [data collection]. 22nd Edition. UK Data Service. SN: 6727, http://doi.org/10.5255/UKDA-SN-6727-23

As our primary outcome, we focus on an indicator for whether a mother is in paid work at the 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[4]). 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 4 by OLS:

$$y^a_{iwat} = \alpha + \delta^a SS_{dq} + \pi_w + \lambda_t + \gamma^{am}_{it} + \phi^{a,k} g^k_{it} * Ki\beta^a X_i + \epsilon^a_{wdmt}, a = 0, ..., 15 \tag{4}$$

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: γ^{am}_{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

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

children k through a continuous measure of the older child's age in years g_{it}^{k}, interacted with an indicator K_i for whether there is such a child in the household.

Unlike our main hospitalization regressions, Equation 4 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 4 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).

D.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 D.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 D.1 then rescales each of these coefficients by the baseline employment rate of women whose youngest child was that age in 1996.

Figure D.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 D.2 to D.4.[5]

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.[6] In Column 2 we control for ward fixed effects and for the set of basic controls shown in Equation 4. In the next three columns we present additional

[5]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.

[6]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.

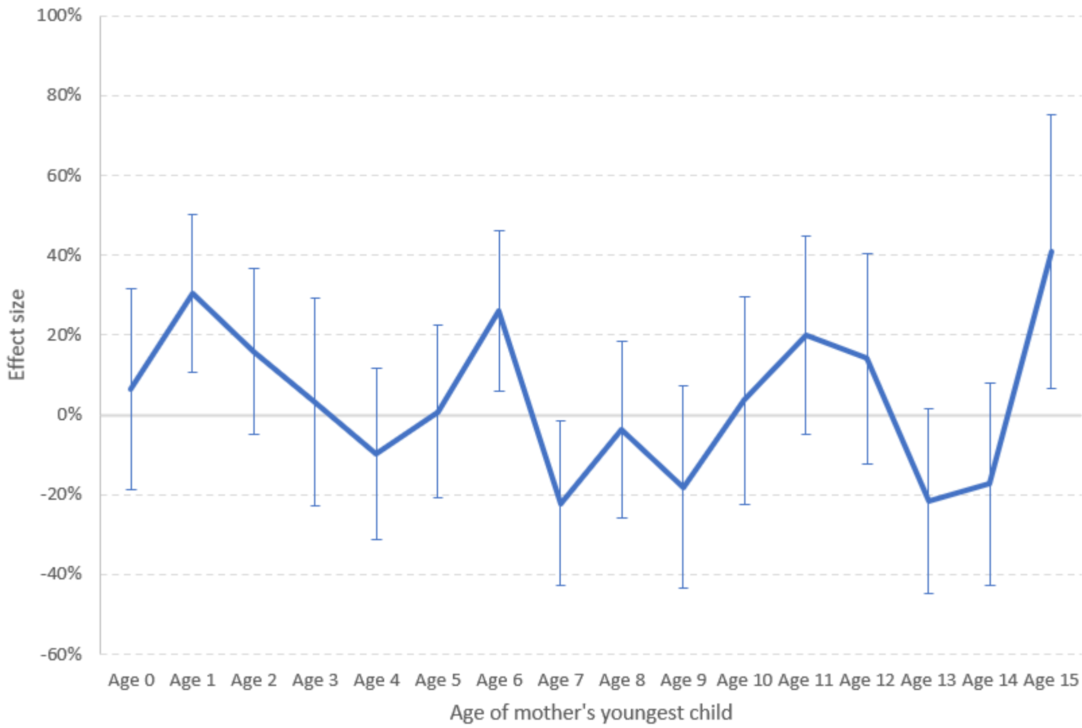
robustness checks. In Column 3 we allow for a local authority-specific linear time trend, estimated based on pre-treatment data and extrapolated to the post-Sure Start period. The inclusion of these estimated trends has very little impact on our results. In Column 4 we additionally control for characteristics 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.[7] 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 D.2). 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.

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

Figure D.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 D.2: 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 D.3: 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 D.4: 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.

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.[8] 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 3.2. 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:

$$D^y_{imda} = \delta^y SS_{dq} + \pi^y_m + \phi^y_q + \psi^y_t + \gamma^y_a + \beta^y X_{it} + \epsilon^y_{imda} \quad (5)$$

[8] Data citation: University of Essex, Institute for Social and Economic Research. (2021). Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009: Secure Access. [data collection]. 12th Edition. UK Data Service. SN: 6676, DOI: 10.5255/UKDA-SN-6676-12.

where D^y_{imda} 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 4, 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 4, 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 4 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 4 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.

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.[9] 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 Table E.5).

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.5: 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.

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

F Cost-Benefit Analysis of Sure Start

This section reports the details of the cost–benefit calculation we report in section 7 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.[10]

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

[10] 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 F.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 F.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. 2024. "Revisiting Event Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies*, 91(6): 3253–3285.

Brewer, Mike, Sarah Cattan, Claire Crawford, and Birgitta Rabe. 2022. "Does more free childcare help parents work more?" *Labour Economics*.

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." Child Accident Prevention Trust.

Conti, Gabriella, Stephen Morris, Mariya Melnychuk, and Elena Pizzo. 2017. "The economic cost of child maltreatment in the UK: a preliminary study." National Society for the Prevention of Cruelty to Children.

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." *Injury Prevention*, 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." *Archives of Disease in Childhood*, 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.

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): 527–532.