WARWICK

ECONOMICS

# Infant Health, Cognitive Performance and Earnings : Evidence from Inception of the Welfare State in Sweden

Sonia Bhalotra, Martin Karlsson, Therese Nilsson & Nina Schwarz

April 2021 No: 1345

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

# Infant Health, Cognitive Performance and Earnings:

## Evidence from Inception of the Welfare State in Sweden

Sonia Bhalotra

University of Essex

Martin Karlsson

CINCH, University of Duisburg-Essen

Therese Nilsson

Lund University, Research Institute of Industrial Economics (IFN)

Nina Schwarz

University of Duisburg-Essen

Acknowledgements: The authors would like to acknowledge the generous support of Riksbankens Jubileumsfond (The Swedish Foundation for Humanities and Social Sciences) P12-0480:1, the Center for Economic Demography, Lund University and the Swedish Research Council dnr 2019-03553. We acknowledge partial funding from ESRC Grant ES/L009153/1 awarded to the Research Centre for Micro-Social Change at ISER, University of Essex. Johanna Ringkvist, Josefin Kilman and Ines Hußmann provided excellent research assistance. We are grateful to Andreas Bergh, Gustav Kjellsson, Alessandro Martinello, Teresa Molina, Cheti Nicoletti, Anton Nilsson, Martin Nordin, Owen O'Donnell, Erik Plug, Peter Sandholt Jensen, Bernhard Schmidpeter and participants of various conferences and seminars for their feedback on earlier versions of this paper.

**Abstract**

We identify earnings impacts of exposure to an infant health intervention in Sweden, using individual linked administrative data to trace potential mechanisms. Leveraging quasi-random variation in eligibility, we estimate that exposure was associated with higher test scores in primary school for boys and girls. However only girls were more likely to score in the top quintile. Subsequent gains, in secondary schooling, employment, and earnings, are restricted to girls. We show that the differential gains for women accrued from both skills and opportunities, expansion of the welfare state having created unprecedented employment opportunities for women.

**Keywords**: Infant health; early life interventions; cognitive skills; education, earnings, occupational choice, programme evaluation; Sweden; gender

**JEL classification**: I15; I18; H41

# 1 Introduction

Spurred by cessation of infant mortality decline, the Swedish government trialled a postnatal intervention from 1 October 1931 to 30 June 1933. It had universal coverage and provided information, support and monitoring of newborn health, including encouragement of breastfeeding, sanitation, a healthy diet and home visiting and clinic attendance. It was a significant pillar in the emergence of the welfare state in Scandinavia. The explicit purpose of the intervention was to bring down infant mortality, important in itself and also as a marker for improvements in infant health for those who survive infancy (Bozzoli et al., 2009). In previous work, we establish that the intervention achieved this goal (leading to a 24% decline in infant mortality) and, in addition, led to meaningful reductions in adult chronic disease mortality and, thereby, to improvements in longevity (Bhalotra et al., 2017).

In this paper we examine dynamic impacts of the improvement in infant health on educational and economic outcomes. Infancy is a period of rapid neurological development – the brain doubles in size in the first year, and by age three it has reached 80% of its adult volume (Nowakowski, 2006). Brain growth is sensitive to nutrition and infection. It is estimated that 85% of calorie intake in infancy is used to build brains, and severe or repeated infections in infancy may divert nutrients away from brain development (Finch and Crimmins, 2004; Eppig et al., 2010). Moreover, the release of inflammatory molecules during infections may directly impact the developing brain by changing the expression of genes involved in the development of neurons and the connections between them (Deverman and Patterson, 2009). Thus there are biological mechanisms for causal effects of infant health on cognition.

The biological mechanisms may be reinforced as follows. Individuals carrying an improved cognitive endowment from infancy may make greater investments in education (lower cost of effort), receive reinforcing investments from parents (Yi et al., 2015; Almond et al., 2017; Bhalotra and Venkataramani, 2013; Adhvaryu and Nyshadham, 2016), and compete more effectively for state investments in education (which, we will argue, played a role in the context we study). If the intervention-eligible cohorts exhibit higher human capital attainment and, if there is sufficient demand for the acquired skills, we may expect them to have higher earnings. We will show that, for our sample cohorts, both skill acquisition and opportunities mattered for

realisation of the impact of infant health on adult labour market outcomes. This is our main contribution. As highlighted in a survey on the long arm of childhood exposures, the evidence on mechanisms or key levers at different points of the life course is particularly scarce (Almond et al., 2017).[1]

We use linked administrative data for a large and representative sample of individuals tracked from birth, through school, to labour market outcomes and then to retirement and death. It is unusual to have individual longitudinal data from birth to death for a population and, especially unusual to have school test scores linked backwards to quasi-experimental variation in birth conditions, as well as forwards to labour market outcomes. We digitised the birth and school records and linked them to available administrative data on later life outcomes. Individual birth certificate data were obtained from historical parish records for 114 rural parishes and 4 cities that we show were representative of the country in 1930. The population of births in these regions is about 25,000 births occurring in 1930–1934. Linkage was done using first name, last name, exact birth date and parish or city of birth. The match rate of births was 66% to school records, 86% to 1970 census files and 65% (91% of survivors) to tax registers. Sample attrition is potentially endogenous because the intervention influenced survival rates but we find it is not differential by treatment status in the census and tax register samples, though it is in the school sample. We nevertheless investigate robustness of all estimates to adjusting for attrition. Match rates were similar for men and women, and attrition adjustments are by gender.

Identification exploits eligibility criteria within treated parishes, and we additionally include matched controls. In treated parishes, children aged 0–12 months at any time in the window for which the programme was available were eligible, for durations that varied with their exact date of birth. We conduct a suite of robustness checks. These include using alternative measures of intervention exposure, investigation of pre-trends, randomisation inference, and sensitivity of the estimates to selective survival.

Our main findings are as follows, and magnitudes are reported for a year of intervention exposure. Primary school test scores at age 10 improved for intervention-eligible boys and girls, although with a markedly different distribution of gains. Treatment effects for boys were similar

[1] A discussion of how our results contribute to specific domains of the literature is in Section 8.

across the distribution (averaging 0.1 standard deviations), while treatment effects for girls were evident only in the upper reaches of the distribution. The intervention increased the chances that girls score in the top GPA quintile by 12.4 percentage points in contrast to an imprecisely determined 2.75 percentage points for boys. In this era, while primary school attendance was universal by mandate (Fredriksson et al., 1971), only about a fifth of all children progressed into secondary school. Secondary school places were limited and, using individual data, we show that the chances of attaining secondary schooling increased sharply for children scoring towards the top of the primary school test score distribution. In line with the intervention having shifted girls into the top GPA quintile, we find that intervention exposure was associated with a significant 3.5 percentage point increase in secondary schooling for girls, alongside no change for boys. To illustrate the role of capacity constraints, we leverage arbitrary variation across secondary school catchment areas in the share of treated children, showing crowd-out that disproportionately hurts boys.

Tracking the intervention and control cohorts over time, we observe an intervention-led divergence between the labour market outcomes of men and women at age 36-40. A year of intervention exposure is associated with an average increase in earnings in the treated population of 7.3%, which is entirely driven by a 19.5% increase in earnings for women. This is large because it reflects an extensive margin increase as well as endogenous increases in skill. As a result of the intervention, women were 5.3% points more likely to participate, they were 7.6% points (20.5%) more likely to be in full-time employment, and almost entirely in the public sector.

Estimates of unconditional quantile treatment effects following Firpo et al. (2009) suggest no gains anywhere in the distribution for men, and that income gains for women were concentrated in the upper part of the distribution. The probability that women belong to the top earnings quintile increased by 8 percentage points. To illuminate this further, we analyse occupational sorting by gender. This reveals that the intervention led to women being 5% points (29.4%) more likely to work as managers and professionals (half of this increase is in the health sector), and 4.4 percentage points (35.5%) more likely to work in accounting, banking and administration. These were among the highest wage occupations in this era. We document

the high skill intensity of these occupations using three measures: the average primary school GPA of workers, the share of secondary school workers and the cognitive skill task content, measured as in Autor et al. (2003).

Using a shift-share approach (Goldsmith-Pinkham et al., 2018; Borusyak et al., 2018), we show significant heterogeneity in intervention effects by an index of the demand for skilled women. The labour market in Sweden was highly gender-segmented at this time. A one standard deviation increase in an index of demand for women's labour at the parish-cohort level almost doubles the estimated impact of the infant intervention on labour market outcomes for women. The lower tail for women and the average results for men suggest that where growth in opportunities was stunted, treatment effects on labour market outcomes were blunted. Although 92% of men as compared with 37% of women were employed full time, it is notable that, in principle, there was room for men to move into more skilled occupations and increase their earnings. This is where it is relevant that the welfare state expanded and increased skilled jobs for women. Health and education were among the top sectors driving the demand for women, together accounting for at least 21% of demand growth. In the years in which our sample cohorts were making decisions about higher education and employment, Sweden was experiencing a rapid expansion of the welfare state which created a disproportionate increase in labour demand in public sector occupations dominated by women.[2]

Mediation analysis is often used to weight the contributions of different endogenous variables to changes in the outcome of interest. Identifying mediators in longitudinal studies of early life interventions is challenging, requiring either additional sources of exogenous variation or strong assumptions regarding the relationship between treatment, mediators and main

[2]We will investigate the role of two other channels. First, we interact the intervention exposure term with parish-level variation in a childcare expansion policy initiated in 1963, and find no evidence that impacts of the infant intervention were increasing in state subsidised childcare. Second, gender norms were evolving in this period and women were returning to work after having children (Datta Gupta et al., 2006; Stanfors, 2003). But, again, we find no heterogeneity in intervention effects by a proxy for parish-level gender norms. A third possibility, that we do not investigate here, is that continuing improvements in child health and survival contributed to liberating women into the labour market, away from replacement fertility and caring for sick children, as was the case in early 20$^{th}$ century America (Bhalotra et al., 2018).

outcomes (cf. Heckman et al., 2013; Heckman and Pinto, 2015; Huber et al., 2017; Dippel et al., 2017). The sequential ignorability condition assumed in some studies is not tenable when, as is common, most outcomes are proxies for human capital. For this reason, it has been customary to report the effects of an intervention on potential mediators alongside effects on the final outcomes of interest, without attempting to weight the contributions of alternative mediators. This is the approach described thus far. We further develop a simple approach to gauge the extent to which the same individuals contribute to the treatment effects on different outcomes. We complement this analysis with the approach of Gelbach (2016) to attribute treatment effects across potential mediators. Without claiming to estimate causal mediator effects, we improve upon the standard descriptive representation by using the natural sequencing of outcomes across the lifecourse

To summarise, we find compelling evidence that an increase in primary school test scores lifted up the potential trajectory among individuals exposed to the infant health treatment. However, improved labour market outcomes were primarily realised only for individuals scoring towards the top of the distribution and, as a result, continuing to secondary school. The results for boys show that capacity constraints can hamper realisation of the full potential of the infant health gain. They also highlight that, under competition, intervention effects on the distribution can determine the size of economic gains. By identifying the skill-content of the occupations that eligible individuals entered, we trace a path from infant health to earnings via skill acquisition. We also demonstrate that the employment and earnings returns to the intervention varied significantly with a measure of growth in the demand for labour. The consistent differentiation of results by gender lends credence to the notion of a causal chain running from earlier to later life outcomes: the higher primary school test scores of women propelled them into higher education, and the growth of skilled occupations in the public sector absorbed them into the labour market. We provide a crude cost-benefit analysis which suggests a very high internal rate of return to the intervention despite the absent results for men.

Our results are relevant to policy design today. Sweden in the 1930s had an infectious disease environment similar to that in many developing countries today. Modern medicine has progressed but there remains considerable scope to improve preventive measures, including

provision of information concerning diet and hygiene, and routine checks or home visits to identify problems that need clinical attention. While universal health coverage, especially for maternal and child health, is high on the current global health agenda (Gorna et al., 2015), there are few systematic evaluations of immediate or long run impacts (Engle et al., 2007). Early childhood programmes similar to the Swedish trial are being introduced in developing countries, for example, the Chilean Crece Contigo Programme (Clarke et al., 2018) and the Indian Integrated Child Development Programme (Dhamija and Gitanjali, 2019), and being refurbished in richer countries, for example the Nurse Family Partnership in the UK (Cattan et al., 2019).

Our findings suggest that a simple low-cost infant health intervention can produce benefits over and above its target, and across domains including infant health, education and earnings. Thus the return to investing in infant health is much higher than is commonly recognized in global health debates. By virtue of showing that an infant health intervention can improve cognitive skills of children, our results also provide new evidence relevant to what is referred to as a global learning crisis, with millions of children failing to attain their cognitive potential (UNESCO, 2014). Knowledge that differences in cognitive skills emerge early and widen with age has led to a call for pre-school interventions (Flavio and Heckman, 2007; Doyle et al., 2009; Attanasio, 2015). Our findings suggest an alternative tool with a similar or larger benefit-cost ratio.

Previous work has discussed changes in the relative demand for female (vs male) labour stemming from recession, war, or technological change (Elsby et al., 2010; Acemoglu et al., 2004; Cortes et al., 2018; Bhalotra et al., 2018). We provide a new perspective, emphasising that expansion of broad-based public services tends to raise the relative demand for female labour. This is of potential relevance to understanding prospects for women in developing countries that are currently witnessing large-scale expansion in the provision of schooling, public health services and pre-school centres. Our findings for men highlight that the earnings payoff to cognitive skills is uncertain, being dependent upon capacity constraints and demand conditions.

The rest of the paper is structured as follows: Sections 2 provides background information

on the intervention and on the educational system in Sweden in the early 20th century. Section 3 describes the data and the empirical strategy, while Section 4 presents the results and Section 5 and 6 discusses potential mechanisms and mediators. Section 7 presents robustness checks. Section 8 delineates our contributions to specific domains of the literature, and also elaborates how our study differs from related studies, and Section 9 concludes.

# 2 Background

## 2.1 The Field Trial – Institutional Details

Following declines in maternal and infant mortality at the beginning of the 20th century, progress stalled during the 1920s, giving rise to an intense public debate in Sweden, and the intervention we analyse emerged as a potential solution. The intervention was described as a trial, implemented prior to a decision on nationwide adoption. It started on 1 October 1931 and ended on 30 June 1933. It was implemented in 7 health districts containing 59 municipalities (2 cities and 57 parishes), namely Lidköping, Hälsingborg, Harad, Råneå, Jokkmokk, Pajala and Mörtfors. They were chosen to be representative of the country in population density and living standards and the selection of districts was not based upon infant or maternal mortality rates, the primary targets of the intervention. The trial was fully funded by the central Government to the tune of SEK 41,400 (USD 139,000 in current prices)(Swedish Government, 1931; SOU, 1935).

To ensure uniform standards of care across the districts, a five-day long educational event for participating staff was organised in Stockholm in July 1931. The trial activities were decentralised to the district level and led by physicians. In each of the seven districts a health centre with regular office hours 2–3 times per week was started. Outreach activities included announcements in local newspapers and churches, and oral announcements by midwives and nurses (Stenhoff, 1934). In total about 2,000 mothers and 2,600 children enrolled.[3] We digitised records maintained by health professionals that indicate programme utilisation. These

[3]On average 72 per cent of all eligible children in rural areas enrolled in the trial. The share of all eligible children that enrolled in the two urban areas was 52 per cent in Lidköping and 32 per cent in Hälsingborg.

show that the average infant made 2.8 visits to a health centre and received 3.9 home visits.

The intervention focused on preventive care and included check-ups at surgeries, home visits and information campaigns. Newborn children were weighed and checked, and sick children were referred to doctors. Mothers were encouraged to breastfeed and given written and illustrated details on the nutritional needs of children at different stages of development. Home visits by nurses were designed to provide advice on hygiene, sanitation and cleanliness in the household, and to ensure that families followed guidelines published by the National Board of Health. So as to understand what the control group in our analysis received, it is relevant to note that while Sweden had a fairly developed primary care system in 1930, there were limited preventive care and support activities targeting infants and expecting mothers.[4]

Eligibility for the infant care programme was determined by birth date. All children less than 12 months of age at the start of the intervention were eligible and eligibility ceased on their first birthday. Appendix Figure A1 shows the duration of eligibility in months for the infant intervention by birth date. An antenatal care programme was introduced simultaneously with the postnatal program and all expectant mothers were eligible, irrespective of their stage of pregnancy. The raw correlation of duration of eligibility for the antenatal and the postnatal interventions is 0.32 and, conditional on eligibility for any one intervention, this falls to 0.13. Given the differential exposure of each individual to the antenatal vs the postnatal components of the programme, we can estimate impacts of each conditional on the other. Our estimating equations consistently include a term measuring exposure to the antenatal care programme. However, in Bhalotra et al. (2017), we found no impacts of the antenatal care programme on infant mortality or the later-life health of the children and, anticipating the results in this paper, we again find no positive impacts of the antenatal program on the economic outcomes of the

[4]The philanthropic childcare institution the *milk drop central* was at the time established in 22 larger cities (among which one city, Hälsingborg, was part of the trial) and engaged in activities to distribute cow milk mixtures to disadvantaged mothers and mothers that could not breastfeed (Wallgren, 1936). The milk drops were generally open twice a week. A Government report of 1929 suggest that they covered around 20 per cent of the infants in the cities where they were established (SOU, 1929). A second type of institution was neonatal care units performing health check-ups and monitoring of mothers and babies. By 1929 there were three active units: in the cities of Stockholm, Gothenburg and Karlskrona (SOU, 1929), cities not included in the trial we analyse.

births. For this reason, the discussion focuses upon the postnatal (infant care) programme.

Annual audit reports in the early 1930s that we perused in libraries indicate programme fidelity, and harmonisation of activities across the treated districts. The trial received positive evaluations from involved physicians in a final report in 1933, attributing improvements in infant health to behavioural change among mothers (Stenhoff, 1934). The first systematic evaluation of the trial is in Bhalotra et al. (2017), where we show that the average duration of programme exposure in infancy led to a 1.56 percentage point decline in the risk of infant death (24% of baseline risk) and a 2.56 percentage point decline in the risk of dying by age 75 (7.0% of baseline risk). We present evidence that intervention-led declines in the risk of dying after the age of 50 were dominated by reductions in mortality from cancer, cardiovascular disease and infections.

## 2.2 The Swedish School System

In the 1930s, schooling in Sweden started in the year an individual turned seven and was compulsory for six years, and primary education (*Folkskolan*) was universal.[5] Sweden had a tracking system whereby students progressing to secondary schools left *Folkskolan* either after grade 4 or after grade 6. On average barely 20 per cent of children attended secondary school, availability of which increased in the 1940s.[6] Importantly, the share of girls and boys attending secondary school was similar in 1930. A reform implemented in 1927 granted equal access for girls to all state-led grammar schools, mandating that girls study the same curriculum as boys.[7]

Teachers in primary school kept records of test scores and attendance in catalogues, which we digitised. The government established several marking principles (see Appendix B). For example guidelines dictated that teachers should reward the quality of knowledge and not

[5]Parents were legally obliged to send their children to school (§51 of the royal decree of the Folkskola).

[6]In the early twentieth century the school system was highly selective. A series of reforms 1925–1945, driven by demand and a political will to reduce educational inequalities between urban and rural areas, increased access and the geographical spread of secondary schools, see Lindgren et al. (2014); Stanfors (2003).

[7]Before 1927 girls could take on secondary education, but only in private schools, the higher costs of which led to lower girl enrollment. By the time our sample cohorts were of secondary school age, the situation was transformed, reflecting rapid increases in girls' attendance, particularly in state schools. A more comprehensive overview of the Swedish school system at the time may be found in Fischer et al. (2019).

the quantity, and take notes throughout the year to ensure that grading reflected performance through the year and not at one point in time. The marks we analyse should thus be purged of day-of-test idiosyncrasies. Teachers were instructed to allow for mark inflation as pupils progressed to higher grades, and to make no adjustment for school form. Thus, the marks reflect an absolute standard and not the relative position of a pupil in their class. The test score data are thus fairly reliable. For our sample cohorts, schooling was fairly comparable across the country[8] and the curriculum did not change between 1919 and 1950. The data contain information on school form, a measure of school quality, and we control for this.[9]

# 3 Data and Empirical Strategy

## 3.1 Administrative Data Linkage

The dataset is unique in linking individual-level data across the life course using birth registers, school registers, the 1970 census and official tax registers. The birth and school registers were digitised by the authors.

**Birth Registers.** A census of 24,390 live births in 1930–1934 was digitised from church records, to include births before, during and after the trial of 1931-1933. Sweden is one of the few countries with high-quality vital statistics at the parish level from the 18$^{th}$ century onwards (Pettersson-Lidbom, 2015). The birth data contain sex, marital status of the mother, age of the mother and parental occupational status, which we translated into occupational classes based on the HISCO classification (Leeuwen et al., 2002) to control for socio-economic status. We merged these birth register data with data from several other sources using linking procedures that were carefully executed and validated;[10] see Bhalotra et al. (2017) for details.

**Administrative School Records.** We accessed standardised exam catalogues containing

[8] A central education plan was introduced in 1919 to overcome differences in the content and format of primary education across Sweden's 2400 school districts. Guidelines published by the Department of Ecclesiastical Affairs included time-tables, syllabi for compulsory schooling, and statement of possible school forms

[9] Appendix Table M1 provides an overview on the proportion of school forms in 1940/1941 in comparison to our sample.

[10] The earliest population census in the lifetime of our subjects that has been digitised is the 1950 census.

pupil-level information from historical archives (see Appendix Figure B1). These contain yearly information on school performance and sickness absence in primary school. We observe the birth cohorts of 1930–1934 in grades 1 and 4 of primary school, in school years 1937–1947. Grades 1 and 4 are pivotal as grade 1 represents the first occasion at which school performance can be observed and grade 4 represents the last as some pupils leave the basic track and proceed to secondary schooling afterwards. The data contain performance in math, writing and reading and speaking, and religion.[11] Other variables include sickness absence and total absence in days, the length of the school year, school type, the name of the teacher and the name of the school.[12] Individuals in the birth records were matched to school records using an algorithm based on birth parish, date of birth, forename and surname. Out of 22,500 individuals still alive at age 7, roughly 16,000 were matched to the school records. It is only for about half of our sample that we have information on both grade 1 and 4, for the rest a child is either observed in grade 1 or in grade 4. Due to the possibility of grade retention there are a few cases where we observe pupils more than once per grade but, at $1.6\%$, grade retention was rare (Hjalmarsson et al., 2015).[13]

**Labour Market Outcomes.** We merged individuals in the birth records to data from the 1970 population and housing census which covers the entire population of Sweden on 1st November 1970 (SCB, 1972). It contains educational attainment, income, employment status and occupation. Of 24,390 births in 1930-34, we observe roughly 20,900 in 1970. Upon matching birth to death registers, we can see that 74% of the 3,490 unmatched individuals died

[11]Religion covered Christianity. There were no detailed curricula in the education plan for this subject, only general information about the course content, and no recommended learning material in addition to the bible (SOU, 1946:11). The focus seems to have been bible reading, and learning of psalms and bible texts.

[12]Sickness absence accounts for about 80% of total absence. Other reasons for absence could be inappropriate clothing or weather conditions preventing children from going to school. For details see (Cattan et al., 2017).

[13]The reason for most of the missing information is that archives of certain schools were accidentally destroyed. Other reasons are death before reaching school age; discrepancies in name spelling; and migration between birth and school age. We significantly reduce the matching problem related to migration by tracking migrants and collecting school records from their destination parishes (not necessarily part of the original sample of treated and matched parishes). Another possible explanation for unmatched individuals could be adoptions since we also match on parental surname, but only 1% of children in our cohorts got adopted (Bernhardtz and Klintfelt, 2007).

before the 1970 census enumeration.[14,15]

**Pension Income.** We linked the birth records to pension (labour) income available for 2001–2005 from official tax registers. These contain information on 16,180 individuals from the birth records (6,621 individuals having died before the year 2002 and 1,589 individuals unmatched). An advantage of using pension income is that it is insensitive to career interruptions such as those associated with childbearing, which could influence income observed in 1970 at a prime working age. For the sample cohorts, obtaining a full pension required thirty years of contributions and the level of the pension was based upon the best fifteen years (Sundén, 2006). Appendix Tables M2 and M3 present descriptive statistics on all explanatory and outcome variables.

**Longitudinal Individual Data: Four points in the lifecycle.** To summarise, after linking the above datasets, we track outcomes at four points in the lifecycle. The potentially treated cohorts are born 1931–1933, and observed in first grade between the school years 1938–1940 when they are 7 years old, and in fourth grade between school years 1941–1943 when they are 10 years old. We then observe them in 1970 when they are age 37–39, a labour market active age. Conditional on survival, we match 72% of the birth sample to school data and 96% of the birth sample to the 1970 census. Appendix Table K13 provides attrition rates by subgroups of eligibility. We observe pension incomes in 2002–2004 when the individuals are 71-73 years old for 91% of survivors (66% of the birth sample). We checked that the match rates were similar for men and women, despite women changing surname at marriage, this being largely because date and parish of birth and first name uniquely identified most people.[16]

**Matched Controls.** Since the intervention took place in seven health districts consisting of 59 municipalities (2 cities and 57 rural parishes), we identified as matched controls, 2 cities

[14]We are consequently left with about 900 individuals who cannot be matched. It is possible that they emigrated.

[15]The earnings information in the 1970 census is regarded to be of high quality, but women who were the partners of a small business owner or a farmer could be recorded as working full-time or part-time while having zero taxable earnings. Since this measurement error might bias our results, we impute incomes of these 2,987 women based on their qualifications and hours worked.

[16]Conditioning on survival, the match rates for men vs women were as follows: School sample: 71% vs 72%, 1970 Census: 96% vs 95%, Pensions: 91% vs 91%.

and 57 rural parishes (belonging to 38 different health districts) using observable parish characteristics from the 1930 census. The best matches (denoted $\mathcal{J}_M(i)$) were identified using the Mahalanobis distance metric; details are in Appendix D, where we also present further tests and descriptive statistics that validate the matches. Summary statistics for a range of relevant observables suggest that our analysis sample is representative of Sweden.[17] Appendix Table D2 shows 1930 census statistics and the standardised difference (Imbens and Woolridge, 2009) between treated districts and the rest of Sweden. It also shows the standardised difference between treated districts and their matched control, indicating balance across groups and validating the matching procedure. To ensure balance among the matching procedure variables, observations from the control group were weighted based on their population size in 1930 relative to the population size of the treated locations they were matched to. On the one hand this reduces potential bias while on the other hand it will slightly reduce the efficiency of our estimates.

## 3.2 Empirical Strategy

We estimate impacts of the infant health intervention on academic performance in primary school, secondary school completion, adult employment, occupation and earnings. We leverage eligibility criteria, whereby children were eligible between birth and the age of 12 months. Importantly, this delivers variation in eligibility within treated parishes. As discussed we additionally use matched controls, so that the estimates are derived from comparing outcomes for exposed cohorts in treated regions to those of unexposed cohorts and control regions. In contrast to the case in most DID designs, our intervention is switched on and off, as a result of which unexposed cohorts include ineligible individuals born before and after the exposed cohorts. We define exposure as duration, using exact date of birth together with the exact dates of the start and end of the intervention. Children born in October 1931 were exposed for the maximum duration of 12 months while children born in January 1932 were exposed for 9 months. In robustness checks we investigate alternative formulations including one that accounts for age of initial exposure. The estimated equation is:

[17] Appendix Figure D1 visualises the sample areas at the municipality (parish and city) level.

$$y_{ipt} = \alpha + \beta T_t + \gamma_p + \tau T_t D_p + \sigma_t + \lambda X + u_{ipt} \tag{1}$$

where $y_{ipt}$ is the outcome for child $i$ born in parish $p$ on day $t$, $T_t$ is the duration of eligibility for the intervention for child $i$ born on day $t$ in years, $D_p$ is a dummy equal to one for treated parishes, $\gamma_p$ are parish fixed effects, $\sigma_t$ are $Quarter\ of\ birth \times Year\ of\ birth$ fixed effects and $X$ is a vector of covariates.

Covariates that we condition on include whether the child was born in a hospital, marital status of the mother, a twin indicator, dummies capturing older (>35 years) and younger (<25) mothers and the occupational status of the household head at the birth of the child.[18] We also control for eligibility for the maternal intervention since some individuals were eligible for both interventions. The richness of the information in the school records allows us to also control for school fixed effects, length of the school year, and school form (an indicator of school quality). In order to allow for differential trends in outcomes between treatment and control regions, we investigate robustness to including parish specific time-trends, which are more general than treatment-group-specific trends. We scrutinize pre-trends in event study plots.[19,20]

The parameter $\tau$ measures the intent-to-treat (ITT) effect of the infant intervention for an additional year of eligibility. This is the parameter of interest for policy makers who are unable or unwilling to make the utilisation of services mandatory. Since there were no always-takers (cf. De Chaisemartin, 2012) the ITT is a scaled version of the average treatment effect on the treated (ATT). As in all studies of the long run effects of a positive health intervention, surviving individuals are negatively selected and, as a result, our estimates will be conservative.

[18] In what follows we present estimates for females and males separately. Results for the full sample (combining females and males) are available in Bhalotra et al. (2019).

[19] We also checked that our findings are robust to including health district fixed effects and health district specific trends. Counties contain health districts which consist of parishes, which are in 99% of cases identical to school districts.

[20] Parish fixed effects and trends will account for trends in outcomes associated with ecological conditions. In particular, goiter is caused by iodine deficits and known to have irreversible effect on brain development, and it was quite common in Sweden in the 1930s. Iodine fortified salt was introduced in 1936, after the end of our intervention. Our empirical strategy accounts for baseline geographical variation in the level of iodine, and the possibly differential impacts of iodized salt on cognitive development across parishes with different initial levels.

In Bhalotra et al. (2017) we presented results which increase our confidence that the programme variation across birth cohort and birth parish that generated the initial improvement in infant health is quasi-experimental. We presented evidence that we could reject the concern of differential pre-trends in infant mortality between treated and control parishes. We showed that estimates using within-mother variation in outcomes are similar, suggesting no selection into programme uptake. Using data that we digitised from practitioner records of programme utilisation, we also found no evidence of selective utilisation, by socioeconomic status or gender (see Appendix E). We presented a test showing no programme impacts on fertility. In fact, Sweden had a law in place 1910–1938 (Lex Hinke), encompassing the trial period, that prohibited spreading information about and advertising contraception (Bygdeman and Lindahl, 1994).

In Section 7, we present a number of specification checks for the long run outcomes. We use alternative definitions of the treatment indicator and test the sensitivity of the choice of marking scale anchoring the grading scale to log of income in adulthood. We show test of balance on a range of baseline covariates and use pre-intervention data to formally test for differential pre-trends, test attrition across treatment status and implement placebo and randomisation inference tests. The sample cohorts were exposed to World War II and to two school reforms. Parliamentary decisions in 1936 and 1937 led to the roll-out of an extension of compulsory school years and of the length of the school year.[21] In Bhalotra et al. (2017) we showed that these reforms are largely unrelated to the intervention studied here, but we nevertheless control for both reforms in our analyses. Sweden was neutral during the Second World War and

[21] All school districts were to have implemented the two reforms by the late 1940s (see Fischer et al., 2019). The term length extension, which extended the school year by 3–5 weeks (8–13%), would affect students in all school years, and the extension of compulsory schooling from 6 to 7 years affected pupils who did not proceed to secondary schooling.

historical sources suggest no educational disruptions for our sample cohorts.[22]

# 4 Results

In this section we presents results for education and earnings, examining test scores (at age 7 and 10), progression to secondary school and earnings (measured when the marginal cohort is 39, and 71). We also explore the proximate sources of changes in these outcomes by examining sickness absence in school, and employment and occupation in adulthood. As a first exercise we create indices of outcomes at the three stages of life and adjust for multiple hypothesis testing.

## 4.1 Effects at Different Ages – Outcome Indices

The many outcomes we will analyse fall into a hierarchy with earnings being the primary endpoint. Nevertheless, the multitude of estimates we present might naturally raise concerns about false discoveries. In order to safeguard against this, we first subject results for multiple outcomes measured at a given age to a multiple hypothesis testing correction. Since many of these outcomes are strongly correlated for substantive reasons (e.g. GPA and secondary schooling completion, or earnings and occupation), we follow Anderson (2008) and construct three indices that take the correlation between outcomes into account: (i) the Age 7 Index includes GPA and top quintile GPA in grade 1, (ii) the Age 10 Index includes top GPA in grade 4 and secondary schooling and (iii) the Adult Index includes top income, log income, log pensions, working full time, municipal employment, federal employment, and high-ranking

[22] In fact the *Folkskola* was one of the main social agents for some 50,000 Finnish children that were evacuated to foster care in Swedish families during World War II. This said, schools were allowed to have shorter breaks in case of limited energy supply, and could cancel regular schooling in case of a threat but any lost days had to be replaced by additional days later on, and in case a teacher was called for military service he had to be replaced by a substitute teacher (Fredriksson et al., 1971). We take care of the latter by controlling for school form and we check whether there are any structural breaks in our school data during the war years. We do not find any evidence of disruption in schooling due to the Second World War.

occupation.[23] Results based on our main specification are in Table 1.

**TABLE 1.** RESULTS FOR OUTCOME INDICES: MULTIPLE TESTING ADJUSTMENT

| | Females | | Males | |
|---|---|---|---|---|
| | N | Estimates | N | Estimates |
| A. Age 7 Index | | | | |
| | 5,830 | 0.0286 | 6,059 | -0.0437 |
| SE | | 0.120 | | 0.084 |
| p val | | 0.811 | | 0.604 |
| BH p val | | 1.000 | | 1.000 |
| B. Age 10 Index | | | | |
| | 10,298 | 0.1319 | 10,617 | -0.0347 |
| SE | | 0.051 | | 0.048 |
| p val | | 0.011 | | 0.473 |
| BH p val | | 0.029 | | 1.000 |
| C. Adult Index | | | | |
| | 10,301 | 0.0764 | 10,619 | -0.0072 |
| SE | | 0.022 | | 0.018 |
| p val | | 0.001 | | 0.694 |
| BH p val | | 0.005 | | 1.000 |

Each coefficient is estimated using specification (1) with local trends. Standard errors are clustered at the parish level. 'p val' present conventional p-values; 'BH p val' presents p-values controlling the false discovery rate following Benjamini et al. (2006).

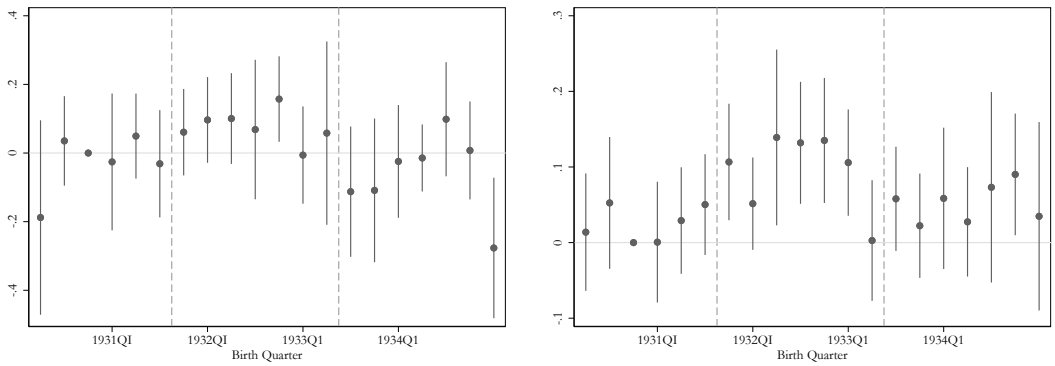
For women, we find statistically significant improvements in educational outcomes of 0.13

[23]Top GPA is a dummy equal to one for an individual in the upper 20% of the GPA distribution, and top income is the corresponding measure for someone in the upper 20% of the income distribution. As discussed in Section 4.2.1intervention-led gains in mean GPA tend to favour boys while intervention-led gains top-GPA tend to favour girls. Since we consistently find significant outcomes differences between the genders, we use top-GPA rather than GPA in the Age 10 Index, the age at which the action is. If we additionally include GPA, the results are noisier but the broad patterns are similar.

standard deviations at age 10, and of 0.07 standard deviations in adult labour market outcomes. For each estimate, we present conventional p values and p values for a multiple testing adjustment controlling the false discovery rate. These are derived using the two-stage procedure proposed by Benjamini et al. (2006).[24] This adjustment leads to higher p values throughout but the results for women remain significant at the same levels as indicated by conventional standard errors.

Figure 1 presents event studies for females for the two indices that were affected by the intervention. There are clearly some power issues but in general, the two event studies confirm the main findings and lend support to our empirical strategy. The rest of this section provides results for each outcome.

**FIGURE 1.** EVENT STUDIES FOR OUTCOME INDICES: FEMALES




(a) Age 10 Index (b) Adult Index

The age 10 index is constructed using top GPA and secondary schooling; the adult index is constructed using top income, log income, log pensions, working full time, municipal employment, federal employment, and high-ranking occupation. Both indices take correlation between variables into account using the method proposed by Anderson (2008). 90 % confidence intervals.

[24] We conduct the adjustment for all six tests at one time; this is different from Anderson (2008) who does it separately by gender. Our adjustment is thus more conservative but given that only two parameters are significant according to conventional p-values, splitting the multiple testing adjustment by gender would not change any result.

## 4.2 Outcomes: Human Capital and Earnings

### 4.2.1 Cognitive Performance - Primary School

Based on the digitised school records we created a measure of cognitive ability by taking the mean of grades in math, reading and speaking and writing to form a grade-point average (GPA), although we shall also report subject-specific estimates.[25] Girls, in general, got better marks than boys, and marks in grade 4 exhibit a higher mean and greater spread than in grade 1.[26] We present results separately for grade 1 and grade 4, and for boys and girls.[27] In order to ease interpretation of the coefficients we transform marks into a $z$ score using the inverse standard normal distribution. Appendix Figure M2 plots these data by gender and grade.

**Quantile treatment effects.** Figure 2 plots unconditional quantile treatment effects, following Firpo et al. (2009), for grade 4 GPA by gender.[28] Boys experienced positive treatment effects across the distribution. For girls, the upper 30% of the distribution is significantly higher on account of the treatment. This gender difference in impacts of the treatment on the score distribution was important for accessing secondary school, see Figure 4 which shows that secondary school completion rates increased sharply with primary school test scores towards the top of the score distribution. Since, on average, 20% of the sample cohorts attained secondary schooling, we estimate regression estimates at the mean and for the probability of scoring in the top quintile of the pooled (male and female) GPA distribution.

**Mean and distribution of test score gains.** Exposure to the intervention leads to a statistically significant increase of about 0.08 standard deviations in average GPA in the full sample. Table 2 shows results by gender. The coefficients are not significantly different by sex, but

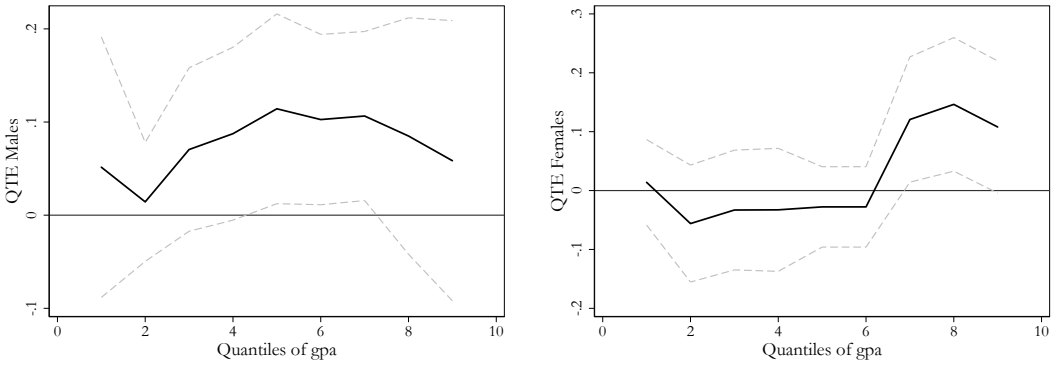
[25]GPA is a common proxy for cognitive performance. When class teachers conduct the assessment, it is possible that test scores also reflect non-cognitive skills. The exam catalogues include marks on behaviour and tidiness, but there is very little variation in these marks, making it hard to analyse them separately.

[26]As discussed in SOU (1942) the recommendation for teachers was to be restrictive with any high or low marks for children in grade 1 and 2, so a lower variation in the first year of Folkskola as compared to grade 4 is expected.

[27]The individual correlation between GPA in grade 1 and grade 4 is 0.46, see Appendix Figure M1.

[28]The correspodning plots for grade 1 GPA show no significant treatment effects, in line with Table 1. Following Firpo et al. (2009), quantiles are defined pre-regression and covariates help adjust for selection bias without redefining the quantiles (see e.g. Borgen (2016) and Killewald and Bearak (2014)).

**FIGURE 2.** QUANTILE REGRESSION: GPA IN GRADE 4 BY GENDER




(a) Boys (b) Girls

Covariates which are included are a dummy indicating twin births, dummies capturing old (>35 years) and young (<20) mothers, a dummy for married women, Parish FE and QOB×YOB FE. 90% confidence intervals included.

larger and only statistically significant among boys, who exhibit a GPA increase of about 0.11 standard deviations (SD). The intervention increases the probability of being in the top quintile by 7.5 percentage points on average, and by 12.4 percentage points for girls in contrast to an imprecisely determined 2.75 percentage points for boys.

At the mean, the significant improvements are in 'writing' and 'reading and speaking', which increase by about 0.11 and 0.12 SD. These increases are not significantly different by gender but larger and only statistically significant for boys. The coefficients for boys are 0.13–0.18 SD, and for girls 0.08–0.11 SD.[29] Appendix Table M4 shows no discernible impact of the programme in grade 1. As some recent studies have found that cognitive gains stemming from pre-school interventions fade (see e.g. Bitler et al., 2016; Chetty et al., 2011), while theory predicts that the gains will multiply over time, it is notable that the infant health intervention

[29]Levine and Schanzenbach (2009) refer to Jacob (2005) to argue that differences in programme impact by subject may arise if performance in some subjects is more sensitive to the value added by school inputs than to other inputs such as the family environment or initial health. Our finding that effects on literacy dominate effects on math is also seen in, for example, Sievertsen and Wüst (2017) who estimate effects of same-day post-birth discharge and Aizer et al. (2018), who estimates effects of reduction in blood lead levels in pre-school, but there are other studies that show similar responses of math scores and reading (see e.g. Figlio et al., 2014; Almond et al., 2014; Bhalotra and Venkataramani, 2013). Thus, we are not alone in finding differences but there appears to be no clear scientific explanation for them.

we consider produced cognitive gains that only become evident at age 10-12. Examining heterogeneity we find that children born out of wedlock benefited substantially more than other children but there were no differences in effects by parental socio-economic status (as indicated by their occupation); see Appendix F.

**Effect sizes – perspective.** To put the average gain in cognitive performance of boys of 0.11 SD in perspective, consider that Bharadwaj et al. (2013) identify effects of 0.15-0.22 SD in Chile and Norway using a sample of children at the low birth weight margin. Using twin fixed effects Bharadwaj et al. (2017) estimate that a 10% increase in birth weight in Chile increases outcomes in math and language by 0.04-0.06 SD, and examining twin pairs in Florida, Figlio et al. (2014) they find the heavier twin scores about 0.05 SD better than the lighter twin. Bhalotra and Venkataramani (2013) find that a 1 SD decline in infant exposure to diarrhea following a water chlorination programme in Mexico led to a roughly 0.1 SD increase in Raven scores and a 0.07 SD increase in math and reading scores. Using the Swedish data used here, Cattan et al. (2017) find that 10 days of sickness absence in primary school led to a reduction in cognitive performance of 0.03 SD. Thus, our estimates are sizeable.

In fact, they look fairly large even in relation to educational interventions in developing countries, some of which have shown test scores gains between 0.17 SD to 0.47 SD (Duflo and Hanna, 2005; Muralidharan and Sundararaman, 2011; Banerjee et al., 2007), while cash transfer programmes have shown limited impacts, with coefficients ranging between 0.04 to 0.08 SD across five studies, and consistently not statistically significant (Baird et al., 2014). Research and policy concerned with improving cognitive attainment has paid increasing attention to the pre-school environment, including parenting styles, caregiver quality and the role of stimulation (Heckman, 2006; Attanasio et al., 2014; World Bank, 2015). Our estimates suggest that pre-school health interventions have the potential to raise cognitive attainment as much as interventions that directly target cognitive capacity.

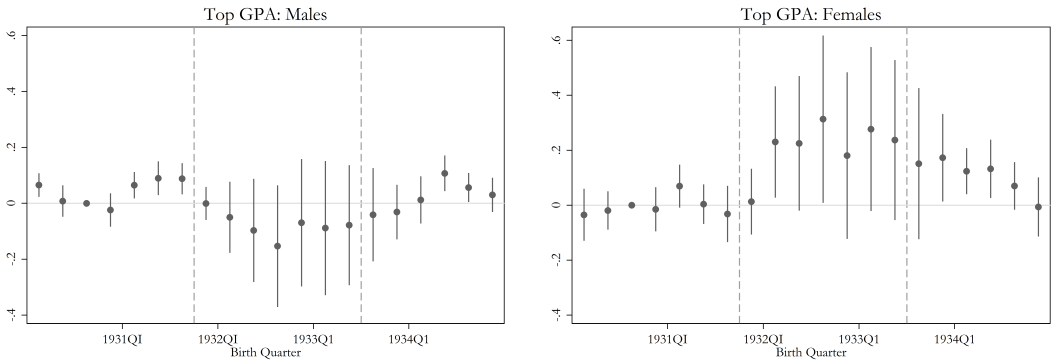
**Event study plot.** Figure 3 presents an event study style plot for the Top GPA outcome, showing coefficient estimates for each quarter of birth. This confirms that improved school performance coincided with treatment eligibility for girls, whereas no such relationship is discernible for boys.

**TABLE 2.** COGNITIVE PERFORMANCE IN PRIMARY SCHOOL, GRADE 4

| | Girls | | | | Boys | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | (1) | (2) | N | Mean | (3) | (4) |
| Top GPA | 6,561 | 0.227 | 0.1000* | 0.1243* | 6,707 | 0.116 | 0.0400 | 0.0275 |
| | | | (0.059) | (0.071) | | | (0.033) | (0.028) |
| GPA | 6,561 | 0.098 | 0.0410 | 0.0617 | 6,707 | -0.200 | 0.1213** | 0.1084 |
| | | | (0.049) | (0.054) | | | (0.057) | (0.072) |
| Math | 6,554 | 0.025 | -0.0535 | -0.0217 | 6,688 | -0.082 | 0.0193 | 0.0317 |
| | | | (0.051) | (0.056) | | | (0.079) | (0.091) |
| Reading | 6,536 | 0.120 | 0.0832 | 0.0902 | 6,687 | -0.241 | 0.1823*** | 0.1649** |
| | | | (0.057) | (0.066) | | | (0.064) | (0.082) |
| Writing | 6,536 | 0.150 | 0.0859 | 0.1068 | 6,692 | -0.275 | 0.1645** | 0.1291* |
| | | | (0.081) | (0.094) | | | (0.064) | (0.072) |
| Religion | 6,549 | 0.088 | 0.0160 | 0.0654 | 6,689 | -0.184 | -0.0222 | 0.0247 |
| | | | (0.052) | (0.066) | | | (0.097) | (0.096) |
| Parish FE | | | ✓ | ✓ | | | ✓ | ✓ |
| QOB×YOB FE | | | ✓ | ✓ | | | ✓ | ✓ |
| School FE | | | ✓ | ✓ | | | ✓ | ✓ |
| SES Effects | | | ✓ | ✓ | | | ✓ | ✓ |
| Length of Schoolyear | | | ✓ | ✓ | | | ✓ | ✓ |
| Schoolform | | | ✓ | ✓ | | | ✓ | ✓ |
| Parish Trends | | | | ✓ | | | | ✓ |

*** p <0,01; ** p <0,05; * p <0,1. Standard errors are clustered at the parish level. Covariates included in all specifications are a dummy for twin births, dummies for old (>35 years) and young (<20) mothers, a dummy for married women, a dummy indicating a hospital birth and the treatment effect of the maternal intervention. 'Mean' refers to the mean value of the outcome variable before the intervention. 'QOB×YOB effects' include quarter-of-birth dummies for each of the 20 quarters. 'Parish FE' are fixed effects for the parish the individual lived in at the time of the birth. 'SES effects' are fixed effects for the professional group of the household head. 'Length of schoolyear' are fixed effects controlling for reforms concerning the length of the school year. 'Schoolform' are fixed effects controlling for the school form as described in Section 2.2 and 'Parish specific linear trends' allows for parish specific time trends.

**FIGURE 3.** EVENT STUDIES: TOP GPA IN GRADE 4 BY GENDER




(a) Boys (b) Girls

The vertical dashed lines signify the eligibility period of the infant care trial. 90% confidence intervals.

### 4.2.2 Secondary Education

Table 3 shows that an additional year of exposure to the intervention resulted in a 3.5 percentage point (17.6%) increase in the probability that girls completed secondary school, while there was no change among boys. The control group mean is somewhat larger for girls, but not significantly different from that for boys. We showed that the intervention led to girls being more likely to score grades in the upper part of the distribution (Figure 2), and baseline performance was already stronger among girls. This is likely to have contributed to the intervention leading to higher secondary schooling increases for girls than for boys – see Figure 4[30]

The greater entry of intervention-eligible girls to secondary school is also consistent with higher returns to secondary school among girls. Using the 1970 census, we regress income in 1970 on test scores in grade 4 and an indicator for completion of secondary schooling (Appendix Table M5). We find higher earnings returns to secondary schooling among girls than

[30]Students seeking entry to secondary education had to take an entrance test (Wallin and Grimlund, 1933). The test was national, covered certain subjects (Swedish and math, written and oral tests) and only students who passed the test were eligible for secondary schooling. For acceptance, students also needed to pass in other subjects in primary school (Dahr, 1945). Despite an increasing number of secondary schools, there were more applicants than available seats, particularly in urban areas. According to Skolöverstyrelsen (1955), about 11 per cent of all applicants of the cohorts born 1930-1934 were rejected. This may contribute to explaining why the intervention did not raise secondary schooling for boys, even though on average they exhibited higher test scores as a result of the intervention. We analyse this carefully below.

**TABLE 3.** SECONDARY SCHOOLING

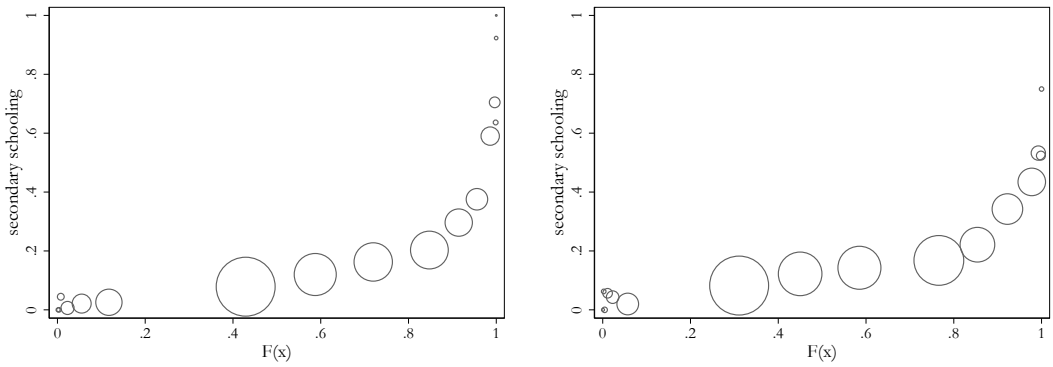
| | Girls (N=10,105) | | | Boys (N=10,369) | | |
|---|---|---|---|---|---|---|
| | Mean | (1) | (2) | Mean | (3) | (4) |
| Primary | 0.675 | -0.0087 | -0.0011 | 0.725 | 0.0280 | 0.0367 |
| | | (0.032) | (0.026) | | (0.023) | (0.025) |
| Dropout | 0.126 | -0.0196 | -0.0277 | 0.101 | 0.0131 | -0.0167 |
| | | (0.023) | (0.023) | | (0.029) | (0.027) |
| Secondary | 0.198 | 0.0353** | 0.0350** | 0.172 | -0.0468 | -0.0289 |
| | | (0.016) | (0.014) | | (0.029) | (0.021) |
| Parish FE | | ✓ | ✓ | | ✓ | ✓ |
| QOB×YOB FE | | ✓ | ✓ | | ✓ | ✓ |
| SES Effects | | ✓ | ✓ | | ✓ | ✓ |
| School Reforms | | ✓ | ✓ | | ✓ | ✓ |
| Parish Trends | | | ✓ | | | ✓ |

*** p <0,01; ** p <0,05; * p <0,1. Standard errors are clustered at the parish level. Covariates included in all specifications are a dummy indicating twin births, dummies capturing old (>35 years) and young (<20) mothers, a dummy for married women, a dummy indicating a hospital birth and the treatment effect of the maternal intervention. 'Mean' refers to the mean value of the outcome variable before the intervention took place. 'QOB×YOB effects' include quarter-of-birth dummies for each of the 20 quarters. 'Parish FE' are fixed effects for the parish the individual lived in at the time of the birth. 'SES effects' are fixed effects for the professional group of the household head. 'School reforms' refers to the extension of compulsory schooling and length of school year reforms, and 'Parish specific linear trends' allows for parish specific time trends.

boys (see also Björklund and Kjellström, 1994).[31] Our findings for secondary school and top quintile GPA are in line with the predictions of Pitt et al. (2012), premised on men having a comparative advantage in brawn-intensive activities, and women in cognition-intensive tasks

[31]Previous work suggests that, in the 1930s too, the returns to years of schooling were greater for women than for men (Bång, 2001) and then lifetime returns to education increased for women in particular following a legal reform implemented in 1939 which prohibited firing women on grounds of marriage or pregnancy, similar to the lifting of marriage bars in the United States (Goldin, 1988).

**FIGURE 4.** CORRELATION OF MARKS IN PRIMARY SCHOOL AND SECONDARY SCHOOLING COMPLETION




(a) Boys (b) Girls

The figure shows mean secondary schooling completion for each possible value of GPA in grade 4 of primary school. Circle size indicates number of individuals in each group.

and later we show occupational sorting by gender consistent with this. Bhalotra and Venkataramani (2013) find broadly similar results in Mexico in the 1990s and Saaritsa and Kaihovaara (2016) in Finland in the early 20th century. Additional explanations that we are unable to test with our data include that the cognitive growth curve differs by gender, that non-cognitive skills such as conscientiousness that are complementary with cognitive skills enhanced girl effort for a given increment to the cognitive endowment, or that the lower labour force participation rates of women in this era led girls to work harder to succeed.

### 4.2.3 Earnings

We estimate that a year of exposure to the infant health intervention raised earnings by 7.3% on average, driven entirely by women experiencing an increase of about 19.5%, in contrast to no gain among men (Table 4). Unconditional quantile treatment effects show no earnings gains anywhere for men but that for women the upper part of the income distribution is moved upward (Figure 5), similar to the pattern observed for cognitive performance. We estimate that the probability of belonging to the top quintile of earners (Top Income) increased by 7 percentage points among women.[32]

[32]We observe that although the cognitive gains apparent for boys did not impact on their earnings, they may have had a positive influence in domains we do not measure, such as financial decision making.

The large earnings increase among women is plausible because (a) as we have seen, eligible women acquired stronger skills and (b) it included an extensive margin increase, which we estimate can account for four-fifths of the observed increase.[33]

**TABLE 4.** EARNINGS

| | Women | | | | Men | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | (1) | (2) | N | Mean | (3) | (4) |
| Top Income 1970 | 10,307 | 0.244 | 0.0655*** | 0.0788*** | 10,613 | 0.210 | -0.0445 | -0.0361 |
| | | | (0.022) | (0.028) | | | (0.034) | (0.028) |
| Log Income | 10,307 | 8.990 | 0.1204* | 0.1947*** | 10,613 | 10.222 | -0.0596 | -0.0464 |
| | | | (0.063) | (0.066) | | | (0.037) | (0.036) |
| Log Pensions (age 71) | 8,284 | 11.609 | 0.0293 | 0.0711*** | 7,680 | 11.995 | -0.0400** | -0.0400* |
| | | | (0.019) | (0.015) | | | (0.017) | (0.020) |
| Parish FE | | | ✓ | ✓ | | | ✓ | ✓ |
| QOB×YOB FE | | | ✓ | ✓ | | | ✓ | ✓ |
| SES Effects | | | ✓ | ✓ | | | ✓ | ✓ |
| School Reforms | | | ✓ | ✓ | | | ✓ | ✓ |
| Parish Trends | | | | ✓ | | | | ✓ |

*** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the parish level. See notes of Table 3 for a list of included covariates, fixed effects and trends.

**Pension income to address measurement error in earnings.** Since earnings are measured at one point in time, in 1970, when the sample cohorts are 37-39 years old, they may be sensitive

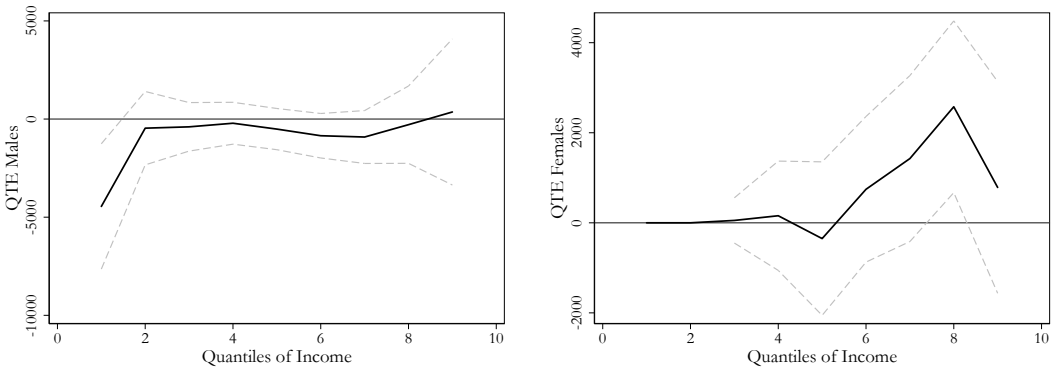
[33]Suppose that prior to the intervention, $n_2$ individuals work full-time, $n_1$ individuals work part-time and $1 - n_1 - n_2$ individuals do not work. Their log earnings are $y_2$, $y_1$ and $y_0$, respectively. After the intervention, $n_2^1$ individuals work full-time and $n_1^1$ individuals work part-time. There were big difference in earnings between women working full-time, part-time (earnings less than half) and not working (earning one seventh; employment refers to the census week and earnings to the year). The extensive margin effect on earnings may then be calculated as

$$\frac{\Delta y}{y_0} = \frac{\left(n_2^1 - n_2\right)\left[\exp\left(y_2\right) - \exp\left(y_0\right)\right] + \left(n_1^1 - n_1\right)\left[\exp\left(y_1\right) - \exp\left(y_0\right)\right]}{n_2 \exp\left(y_2\right) + n_1 \exp\left(y_1\right) + n_0 \exp\left(y_0\right)} \tag{2}$$

In our case, $n_1^1 - n_1 = 0$, $n_2^1 - n_2 = 0.076$, $y_2 = 9.89$, $y_1 = 9.18$, $y_0 = 7.93$. Hence, we get:

$$\frac{\Delta y}{y_0} = \frac{0.076 \cdot 16,953}{8,022} = \frac{1,288}{8,022} = 16\% \tag{3}$$

**FIGURE 5.** QUANTILE REGRESSION OF INCOME BY GENDER




(a) Men (b) Women

Covariates which are included are a dummy indicating twin births, dummies capturing old (>35 years) and young (<20) mothers, a dummy for married women, Parish FE and QOB×YOB FE. 90% Confidence Intervals included.

to lifecycle variation in labour supply, important for women on account of fertility.[34] Pensions mirror the best fifteen years in the labour market and thus represent earnings at advanced stages of the career. Investigating pension income at age 71 as an alternative measure of income, we identify increases in pension income for women of 7%, and no increase for men (Table 4), ratifying the earnings results.[35,36]

**Internal rate of return.** The intervention cost approximately SEK 41,400 (USD 139,000 in current prices). Personnel costs (salaries for physicians and nurses) accounted for 50% of total costs. The cost per treated child was about USD 39 (in current prices) and per consultation USD 5.7 (in current prices). These costs are low relative to the benefits we identify. We calculate the

[34]The cohort fertility rate for women born 1930-34 was about 2.2. As mentioned in the Introduction, we show in Bhalotra et al. (2017) that there is no programme effect on fertility.

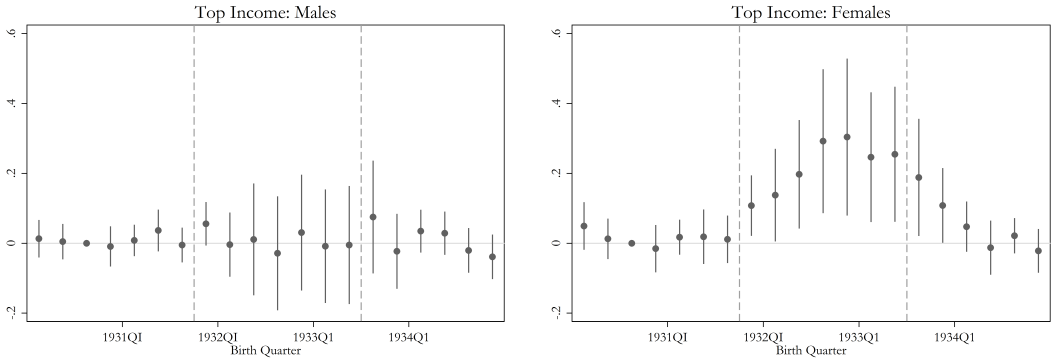
[35]For men, we estimate a decline in pension income of 4%. Since we saw no decline in earnings for men at age 37-39, and since only 63% survive to the age of 75, this may reflect endogenous survival selection, the marginal surviving individual being negatively selected post-intervention (see Bhalotra et al. (2017)). To investigate the role of survival selection, we re-estimated programme effects on 1970 income for subsamples of individuals surviving until age 40, 50, 60, 70 and 75 respectively (Appendix Table K10). We see no selection among females until age 75, when there appears to be some positive selection. In contrast, among men, there appears to be negative selection from age 60 onwards as the earnings estimates become progressively lower the older the age group.

[36]A potential concern with the pension variable is that a widow pension was available to the sample cohorts, and this could create a wedge between women's earnings and their pensions. However the results are robust to controlling for an indicator for whether the individual was in receipt of a widow pension (Appendix Table K18).

net present value of earnings and estimate the internal rates of return on the funds spent by the national government as 0.22, see Appendix L.

**Event study plot.** Figure 6 plots the event study for top income. As seen for top GPA, an increased probability of belonging to the top income quintile coincides with treatment eligibility for women but not for men.

**FIGURE 6.** EVENT STUDIES: TOP INCOME BY GENDER




(a) Men (b) Women

The vertical dashed lines signify the eligibility period of the infant care trial. 90% confidence intervals.

## 4.3 Intermediate Outcomes - Illuminating Mechanisms

### 4.3.1 Sickness Absence in Primary School

There are two channels through which infant health may have had the noted impacts on school performance. First, infant health may predict school-age health, creating a contemporaneous effect from healthy children missing school less often or concentrating better when at school. The second channel operates through brain development and runs directly from infant health to later life cognitive performance (see e.g. Eppig et al., 2010). We investigated intervention effects on sickness absence, as a marker of school-age health, with a view to discriminating between the two channels. Focusing on grade 4, where we saw intervention effects on performance, Appendix Table M6 and Appendix Figure M3 show that the intervention reduced sickness absence for boys by about 0.8% (20%), and it is possible this contributed to their

higher GPAs. However, we see an unexpected increase in sickness absence for girls.[37] Given that we find increases in secondary school and earnings for girls and not boys, this undermines the relevance of the pathway involving morbidity in the school years in favour of the argument that the intervention improved neurological development. We underline that this interpretation is only suggestive.

### 4.3.2 Employment

Women exposed to the intervention for a year exhibited an increase in the propensity to work full-time of 7.6 percentage points, or 20.5%, and no change in the propensity to work part-time (Table 5). More women joined the labour force, this is explicit in the next section.[38] There are no significant impacts on employment for men, 92.5% of whom worked full-time.

**Relative demand for women.** In the years when our sample cohorts were making the relevant decisions, there was a substantial expansion of the welfare state and a sharp increase in the share of working married women (Stanfors, 2003). We posit that these phenomena are related, as the growing welfare state created more jobs for women than for men, as nurses, teachers and child-care workers. Although our estimates exploit a discontinuity in intervention eligibility conditional upon general trends, we argue it was important that intervention-treated individuals emerging on the market with enhanced skills, faced an expansion of job opportunities. In a similar vein, Coles and Francesconi (2017) argue that expanding job opportunities for women was critical to realisation of the impacts of the contraceptive pill on women's outcomes in America. To investigate the role of opportunities, albeit indirectly, we examined the sectors that women responding to the intervention joined, and linked this to historical information on sectoral growth trends. We discuss these results in Section 5.2.

[37] An improvement in health for boys relative to girls is consistent with the stylized fact of boys being more sensitive to health inputs in infancy, although baseline sickness absence rates are similar for boys and girls at about 5% of school days. The distribution and mean of sickness absence for this 1930s births sample resembles closely that in contemporary research (Aucejo and Romano (2014), Goodman (2014)), see Cattan et al. (2017) for further analysis for our sample.)

[38] Part-time refers to 20-35 hours per week and full-time work to more than 35 hours. Both are thought to be underestimated in the 1970 population and household census (cf. Population and Housing Census 1970, 1972), but this applies to men and women and the under-estimation is unlikely to be correlated with the infant intervention.

**TABLE 5.** EMPLOYMENT

| | Women (N=10,256) | | | Men (N=10,466) | | |
|---|---|---|---|---|---|---|
| | Mean | (1) | (2) | Mean | (3) | (4) |
| Working Parttime | 0.265 | -0.0325 | -0.0244 | 0.019 | -0.0077 | -0.0049 |
| | | (0.030) | (0.033) | | (0.007) | (0.007) |
| Working Fulltime | 0.370 | 0.0607* | 0.0760** | 0.925 | -0.0052 | -0.0061 |
| | | (0.031) | (0.037) | | (0.014) | (0.015) |
| Municipal | 0.238 | 0.0377* | 0.0488** | 0.092 | 0.0012 | 0.0102 |
| | | (0.020) | (0.020) | | (0.014) | (0.016) |
| Federal | 0.051 | 0.0306*** | 0.0339** | 0.111 | -0.0053 | -0.0077 |
| | | (0.012) | (0.014) | | (0.019) | (0.019) |
| Parish FE | | ✓ | ✓ | | ✓ | ✓ |
| QOB×YOB FE | | ✓ | ✓ | | ✓ | ✓ |
| SES Effects | | ✓ | ✓ | | ✓ | ✓ |
| School Reforms | | ✓ | ✓ | | ✓ | ✓ |
| Parish Trends | | ✓ | | | ✓ | |

*** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the parish level. See notes of Table 3 for a list of included covariates, fixed effects and trends.

### 4.3.3 Occupation and Skill

**Public Sector Jobs.** Using indicators for employment in municipal and federal government employment, we find that eligibility for the intervention for a year associated with an increase in the probability that women work in municipal jobs of 4.9 percentage points, or 20.5% relative to the baseline of about 24% and an increase in the probability of working in federal governmental jobs of 3.4 percentage points, or 66.5% (Table 5). Adding up across both categories, it appears that more or less all of the additional employment of women was in the public sector. This lines up with the rapid growth of the Swedish welfare state from the mid-20$^{th}$ century absorbing women (Stanfors, 2003; Datta Gupta et al., 2006). Appendix Figure M4 shows how female employment rapidly increased from about 800,000 employed women in 1950 to

about 1,200,000 in 1970, while male employment stayed fairly constant over time.[39] Appendix Figure M5 illustrates the trend in women working in selected public sector jobs 1950–1975.

**Occupation.** So as to more clearly depict the destinations of women, and to make explicit the skill-content of their tasks, we examined treatment effects on occupation. We find that increases in women's employment were concentrated in high-skilled sectors (Table 6). Women exposed to the intervention for a year were 5.0 percentage points (29.4%) more likely to work as managers and professionals and 4.4 percentage points (35.5%) more likely to work in accounting, banking and administration (almost all of this increase is as office workers or administrators). In contrast, we see a reduction in the share of men in the 'managers, professionals' category and an increase in the share of men in sales. The mean earnings by occupation confirm that the highest-paying occupational group was 'managers, professionals', so these findings line up with our finding that treatment led to women being more likely to appear in the top quintile of the earnings distribution.[40] Disaggregating the occupational categories that attracted women further, we find that the largest increase in this category comes from women working in the health sector, for instance as midwifes or nurses (Appendix Table M7). Table 6 also shows, consistent with the employment estimates and our finding of extensive margin changes, a reduction in the *out of the labour force* group for women but not for men.

[39]Notably, the increasing female labour force participation from the early 1950's in Sweden is in quite sharp contrast to the corresponding development in Norway. After the World War II until 1960, the share of Norwegian women in the workforce declined. In 1970 only about 25 percent of married women in Norway were employed. The different development compared to Sweden has been related to higher fertility, but also that the welfare state expansion took place later (Egg-Henriksen, 2008). The universal welfare state was only established in Norway in 1967 (Bütikofer et al., 2019).

[40]Mining and Crafts shows a higher return for women though not for men. We disregard this aberration as 0.1% of women are in mining and 0.6% in crafts.

**TABLE 6.** OCCUPATIONAL SORTING

| | Women (N=10,301) | | | | Men (N=10,619) | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | | | | Mean | | | |
| | Outc. | Earn. | (1) | (2) | Outc. | Earn. | (3) | (4) |
| A. Managers, | 0.176 | 23,909 | 0.0427** | 0.0495*** | 0.224 | 44,196 | -0.0229 | -0.0373** |
| Professionals | | | (0.019) | (0.019) | | | (0.021) | (0.019) |
| B. Accounting, | 0.124 | 18,825 | 0.0388 | 0.0443* | 0.036 | 32,997 | -0.0141 | -0.0210 |
| Administration | | | (0.027) | (0.025) | | | (0.016) | (0.017) |
| C. Sales | 0.083 | 13,063 | -0.0245 | -0.0226 | 0.083 | 33,742 | -0.0052 | 0.0191* |
| | | | (0.018) | (0.017) | | | (0.014) | (0.011) |
| D. Agricultural | 0.026 | 3,260 | 0.0099 | 0.0070 | 0.093 | 21,976 | 0.0081 | 0.0085 |
| | | | (0.007) | (0.007) | | | (0.012) | (0.014) |
| E. Mining | 0.001 | 24,678 | 0.0007 | 0.0003 | 0.036 | 29,266 | 0.0047 | 0.0024 |
| | | | (0.001) | (0.001) | | | (0.008) | (0.009) |
| F. Transport, | 0.031 | 17,346 | -0.0081 | -0.0062 | 0.079 | 27,522 | -0.0002 | 0.0141 |
| Communication | | | (0.012) | (0.011) | | | (0.013) | (0.015) |
| G. Crafts | 0.006 | 31,335 | -0.0206 | -0.0161 | 0.335 | 26,632 | -0.0131 | -0.0286 |
| | | | (0.019) | (0.018) | | | (0.020) | (0.021) |
| H. Service | 0.130 | 11,288 | -0.0087 | -0.0033 | 0.041 | 29,953 | 0.0278 | 0.0238 |
| | | | (0.015) | (0.016) | | | (0.019) | (0.020) |
| I. Out of LF | 0.370 | 2,282 | -0.0301 | -0.0528** | 0.072 | 9,665 | 0.0149 | 0.0190 |
| | | | (0.024) | (0.026) | | | (0.014) | (0.015) |
| Parish FE | | | ✓ | ✓ | | | ✓ | ✓ |
| QOB×YOB FE | | | ✓ | ✓ | | | ✓ | ✓ |
| SES Effects | | | ✓ | ✓ | | | ✓ | ✓ |
| School Reforms | | | ✓ | ✓ | | | ✓ | ✓ |
| Parish Trends | | | | ✓ | | | | ✓ |

*** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the parish level. We provide means of the dependent variables as shares of men and women working in the occupational category at baseline (Outc) and mean earnings for each occupation (Earn). See notes of Table 3 for a list of included covariates, fixed effects and trends.

Table 6 showed some evidence of women displacing men. To investigate this more thoroughly, we went down to the 3-digit level to identify the highest-earning occupations for each gender. We divided the sample into individuals with earnings in the top-20 and the remaining 80% of the distribution. We then ranked 3-digit occupations by their employment shares, retaining the eight occupations that employed half of all individuals in the top-quintile. See Appendix Table H7. First, there is close correspondence between top quintile earnings and public sector employment for females but not for males. Second, the labour markets for males and females were clearly segregated at this time – the top-ranked occupations for women and men exhibit limited overlap.[41]

# 5 Mechanisms – Skill Acquisition and Opportunities

The identified gender differences in programme impacts on labour market outcomes are not a trace of gender differences in the initial programme impacts, as the programme had statistically indistinguishable impacts on infant mortality among girls and boys, or gender differences in utilisation (Bhalotra et al., 2017). We have argued that the sharp differentiation of labour market outcomes between men and women can be explained by (i) treatment having led to greater skill acquisition among women than men, and (ii) stronger growth in labour demand for treated women. This section investigates these mechanisms further.

## 5.1 Skill Acquisition

Boys exposed to the infant intervention registered fairly large and statistically significant improvements in school performance at age 10, and yet their labour market outcomes did not improve. We have argued that the reason that primary school performance translated into long-term gains for girls is that they were more likely to appear in the upper regions of the test score

[41] Appendix Table H8 shows the share of men and women in 1970 in each of the main 2-digit occupations, their mean earnings and skill intensity. Skill intensity is shown using three independent measures: average GPA in the occupation, share of workers with secondary education and the average task content classified as routine vs non-routine cognitive vs non-cognitive following Autor et al. (2003). The two highest-ranked occupational groups ('Managers & Professionals' and 'Accounting, administrative') are high-skilled by all three criteria.

distribution, which made them competitive for entry to rationed secondary schooling places. Below we demonstrate that rationing of secondary school places disproportionately hurt boys. An alternative explanation would be that treated boys faced higher opportunity costs. However Fischer et al. (2019, Appendix F) show that only a very small fraction of each cohort entered regular employment right after finishing compulsory schooling. Therefore, gender differences in opportunity costs are unlikely to have mattered much.

In the early 1940s, when the subjects of our study took decisions on whether to attend secondary schooling, every parish/city (municipality) had a primary school but only 194 of a total of 2,500 municipalities had a lower secondary school (Lindgren et al., 2019).[42] The share of each cohort that took secondary schooling varies between 7 and 42 per cent at the parish level, and most of the variation is between catchment areas. The infant intervention was delivered at the level of a health district. There were 400 health districts and thus, importantly, fewer secondary school locations than health districts. As a result, the competition faced by a treated child in gaining entry to secondary school varied arbitrarily as a function of the share of treated children in the catchment area of the secondary school closest to them (any secondary school being fed by both treated and untreated locations). This is compounded by variation in cohort size and birth date. Figure G1 plots the identifying variation- the share of treated children in the catchment area of the nearest secondary school for a treated child, and Figure G2 confirms that children in the control group were not exposed to treated children.

We exploit this arbitrary variation in order to test the crowding-out mechanism. For each individual $i$, we measure exposure to other treated children as

$$Exp_i = \frac{\sum_{j:p_j \in CA_{S_{p_i t_i}}} T_{t_j} D_{p_j} 1\left(d_{p_j s_{p_i t_i}} < 50\right)}{\sum_{j:p_j \in CA_{S_{p_i t_i}}} 1\left(d_{p_j s_{p_i t_i}} < 50\right)} \tag{4}$$

where $p_i$ and $t_i$ represent the birth parish and birth date of individual $i$, $S_{p_i t_i}$ is the nearest secondary school from parish $p_i$ at date $t_i$, and $CA_{S_{p_i t_i}}$ is the catchment area of that school for birth date $t_i$.

[42]The number of secondary schools did not respond to the treatment. A municipality is typically a rural parish or a city. According to the 1930 census, 16 rural municipalities included several church parishes, and 38 rural parishes included several municipalities (Statistics Sweden, 1937).

We assign each parish to the catchment area of the closest secondary school, with distance calculated as the great-circle distance between parish centroids. $T_{t_i}$ represents duration of eligibility and $D_{p_I}$ is a dummy taking on the value 1 if parish $p_i$ was treated. Since most pupils in a secondary school came from within a radius of 50 kilometres, we calculate exposure imposing this restriction.[43]

Figure 7 presents estimates showing how the treatment effect on secondary schooling completion varies with the share of treated children across different thresholds. The horizontal axis represents percentiles, n, of the distribution. The coefficients show treatment effects by gender on secondary schooling when the share of treated children is at least n. The figure provides evidence that treatment led to crowd-out and, moreover, that boys suffered crowd-out more than girls. Once the share of treated children crosses 20%, there is a significant decline in the probability that a treated boy completes secondary school. A treated girl also suffers competition but the decline in probability for girls is smaller and only statistically significant once the share of treated girls exceeds 65%.[44] Appendix Table G3 shows that this rationing of secondary school places is mirrored in male labour market outcomes in the long term. We investigate labour market outcomes from a different angle in Section 5.2, highlighting that while there was possibly some displacement, it was not a zero sum game at the labour market level as labour demand was expanding overall, albeit more favourably for women.

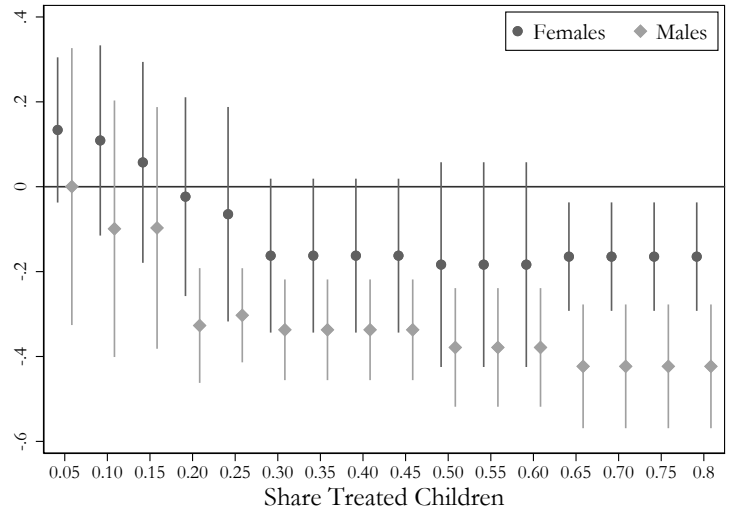
## 5.2 Growth in Labour Market Opportunities

Below we show evidence that (a) there was stronger employment growth in sectors dominated by women, (b) long run effects of the infant intervention were increasing in parish-level em-

[43] Every individual in the dataset gets assigned to their nearest secondary school, regardless of whether they live within the 50 km radius or not. We impose the 50 km radius on the measure of competition from treated children – so that treated children outside the radius are disregarded. This is based on the fact that secondary schooling take-up was negligible outside this radius. Still, $Exp_i$ is the relevant measure of competition also for individuals living outside this radius.

[44] Appendix Table G2 shows estimates collapsed at 50%. Our results may be confounded by effect heterogeneity if it is correlated with the share of treated children. However such heterogeneity is unlikely to give rise to the monotonous relationship we observe for both genders in the figure.

**FIGURE 7.** SHARE OF TREATED CHILDREN: SECONDARY SCHOOLING




Each dot refers to treatment effects on secondary schooling completion when the share of treated children in the catchment area is at least the share of treated children noted on the x-axis. 90% confidence intervals.

ployment growth but only for women, and (c) this result is driven by skilled women, in line with our earlier results showing that treatment raised skills and led to women entering high-skilled occupations. (d) We also describe, using 3-digit occupations, occupational sorting by gender, demonstrating gender segmentation and showing that women were much more likely than men to be in high skilled public sector employment.[45]

Table H7 shows the three-digit occupations of men and women with earnings in the top income quintile and in the rest of the distribution in 1970, by gender, demonstrating that the labour market was gender segmented. This exercise also demonstrates that women were significantly more likely to be working in the public sector, and it confirms that the public sector occupations that women were drawn into were skilled high-wage occupations. Three of the eight most populous occupations in the top quintile of the earnings distribution for women were teacher, nurse and medical assistant. In fact teachers earn in the top quintile of the male earnings distribution, see Figure H4. Women in these three occupations accounted for a third of all women in these largest eight occupations in the top earnings quintile, in contrast to which only 6% of men were public sector workers (teachers). The occupations that dominate the top earnings quintile for men are engineers and architects. Understand sectoral compositional

[45]Recall that higher education or skill for these cohorts amounted to secondary schooling.

change is thus relevant to understanding the growth in women's employment at the top of the distribution.

We leverage plausibly exogenous variation across parishes in baseline industrial composition to predict the growth of demand for female and male labour using a Bartik shift-share approach (Goldsmith-Pinkham et al., 2018).[46] We use 3-digit industry shares in 1950 in the parish of birth and national employment growth 1950–70 in each industry by age and sex, and consider two versions of the index; one for workers in general and one for skilled workers (with more than primary schooling). Our focus is on the latter. The data and procedure are detailed in Appendix H, where we also plot the distribution of the Bartik index. Appendix Table H9 shows that the index is predictive of employment in 1970 after controlling for parish of birth and cohort fixed effects. Among skilled workers, a 1 standard deviation increase in the Bartik index for skilled workers is associated with an 18 percentage point increase in the share of women in work and a 3 percentage point increase in the share of men, and both changes are statistically significant. A reason for the difference is that, at baseline, 97% of men and only 74% of women in the skilled group were in (full or part time) employment. Women responded to demand on the extensive margin, whereas demand growth is more likely to have influenced hours of work or wages for men. The table shows that the skilled Bartik for men predicts an increase in the probability that male earnings are in the top quintile.

Results for the index of adult outcomes are in Table 7. In columns (1) and (3) we interact the treatment term describing exposure to the infant intervention with the Bartik index (for skilled workers) of own gender, and in columns (2) and (4) we additionally include an interaction with the Bartik index (for skilled workers) of the opposite gender. The latter allows us to interpret the own-gender interaction as conditional on general labour market conditions. It also tests for any spillover effects of employment growth in, for example, a female-dominated sector on male outcomes. We transform all the Bartik indices into $z$ scores, so that the estimate for *Treated* $\times$ *Duration Eligibility* represents the treatment effect at the mean of the Bartik index.

[46] As shown by Borusyak et al. (2018), it is not necessary to assume that the original industry shares are exogenous to achieve identification. Assuming instead that the shocks affecting different industries are exogenous also leads to identification. In our case, the identifying assumption would be that conditional on parish fixed effects and cohort trends, the industry-specific shocks are random.

The estimates show that treatment effects associated with the infant intervention were are increasing in predicted labour demand for skilled female workers at the parish-cohort level. This is not the case, on average, for men although there are meaningful positive effects for men born in parishes with very large values of the Bartik index. There is considerable heterogeneity among women. Increasing the Bartik index by one standard deviation almost doubles the effect of the infant intervention on long-run outcomes for women and, conversely, women born in parishes in the lower tail of the Bartik index experienced no positive gains from eligibility for the infant intervention. In Appendix Table H10 we show that this same pattern of heterogeneity holds across the different labour market outcomes we consider, although with varying levels of statistical significance.

The interaction term involving the opposite gender Bartik is not significant– women's labour market outcomes are not sensitive to male employment growth at the parish level, and vice versa. This is consistent with gender segmentation of the labour market at this time. Repeating the analysis using indices for all workers rather than only skilled workers, there is no longer any evidence that labour demand is a significant moderator of the effects of the infant intervention (Appendix Table H11). This is consistent with the intervention having led to higher skill acquisition. It happened to be a time when there was growth in demand for skilled workers (and, as showed before, this happened to be stronger in woman-dominated occupations in the public sector). Rotemberg weights computed by decomposing the Bartik estimator into a weighted sum of the just-identified instrumental variable estimators that use each industry share as a separate instrument reveal that the health care and education sectors were in the top-5 sectors predicting demand for skilled women, accounting for at least 21% of the growth in demand, see Appendix Table H13.

Overall, our findings suggest that the average earnings payoff to infant health was uncertain, being dependent upon capacity in higher education and on context-dependent labour demand conditions.

**TABLE 7.** TREATMENT EFFECT HETEROGENEITY BY BARTIK INSTRUMENT FOR SKILLED WORKERS, ADULT INDEX

| | Females (N=10,301) | | Males (N=10,619) | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treated × Duration Eligibility | 0.0721*** | 0.0747*** | -0.0147 | -0.0142 |
| | (0.022) | (0.021) | (0.016) | (0.016) |
| Treated × Own Skilled Bartik | 0.0070 | 0.0030 | 0.0366** | 0.0366** |
| | (0.051) | (0.051) | (0.018) | (0.018) |
| Own Skilled Bartik | 0.0409 | 0.0453 | -0.0003 | 0.0002 |
| | (0.039) | (0.038) | (0.009) | (0.009) |
| Duration Eligibility × Own Skilled Bartik | -0.0309** | -0.0311** | -0.0216* | -0.0224* |
| | (0.013) | (0.013) | (0.012) | (0.012) |
| Treated × Duration Eligibility × Own Skilled Bartik | 0.0581*** | 0.0587*** | 0.0169 | 0.0196 |
| | (0.018) | (0.018) | (0.017) | (0.019) |
| Treated × Other Skilled Bartik | | 0.0229 | | 0.0352 |
| | | (0.017) | | (0.040) |
| Other Skilled Bartik | | -0.0321*** | | -0.0178 |
| | | (0.010) | | (0.021) |
| Duration Eligibility × Other Skilled Bartik | | 0.0161 | | 0.0014 |
| | | (0.014) | | (0.009) |
| Treated × Duration Eligibility × Other Skilled Bartik | | -0.0105 | | -0.0015 |
| | | (0.019) | | (0.012) |
| Parish FE | ✓ | ✓ | ✓ | ✓ |
| QOB×YOB FE | ✓ | ✓ | ✓ | ✓ |
| SES Effects | ✓ | ✓ | ✓ | ✓ |
| School Reforms | ✓ | ✓ | ✓ | ✓ |
| Parish Trends | ✓ | ✓ | ✓ | ✓ |

*** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the parish level. The Bartik index is defined at the cohort-parish level for skilled workers for each gender. 'Own' is own-gender 'Other' is other gender, see Appendix H for definitions. See notes of Table 3 for a list of included covariates, fixed effects and trends.

### 5.2.1 Were Gender Norms Important in Driving Outcomes for Women?

Expanding opportunities for women, in part related to the expansion of the welfare state, seems to be an important part of the story. We now investigate the potentially competing or complementary hypothesis that regional differences in gender norms contributed. Identification is difficult as gender norms are only ever measured by partial indicators and they tend to evolve slowly, making it is difficult to get credible variation within parishes. We rely on baseline heterogeneity between parishes in female representation in local councils (our proxy for gender equality norms). Beaman et al. (2012) show that gender quotas in local government lead to higher aspirations and educational attainment among girls in India, demonstrating that local women leaders can have tangible effects on women's economic participation.

During the first decades of the 20$^{th}$ century, Sweden made a transition from predominantly direct democracy to representative democracy at the local level (Hinnerich and Pettersson-Lidbom, 2014). Statistics Sweden published data on election outcomes at the municipality level (cf. Statistics Sweden, 1947), including the number of elected women and the total number of seats in each local council. We digitise and use data for the 1946 election as this is the first local election in which a sufficient number of municipalities were practising representative democracy, thus generating sufficient variation in the share of elected women. We define an indicator for above-median female representation in the birth parish. The median share of women in the local council is 0.07, and 20 per cent of our individuals were exposed to zero female representatives.

Since the 1946 election took place 13 years after the end of the infant intervention, we may be concerned that women's political representation is an outcome of the intervention. However, we are able to leverage identifying variation within the treated group using differences in eligibility within parish by birth date, and then interact this with differences in female representation between treated parishes. The estimates for the Adult Index are in Table 8, and estimates for the range of component labour market outcomes are in Appendix Table G4. The coefficients on the interaction terms are small and insignificant. We thus conclude that local gender norms, at least as captured by women's local political power, did not moderate the translation of infant health improvements into improved labour market outcomes for women.

**TABLE 8.** EFFECT HETEROGENEITY: FEMALE REPRESENTATION, ADULT INDEX

| | Females (N=10,301) | | Males (N=10,619) | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treated × Duration Eligibility | 0.0589** | 0.0753*** | -0.0148 | -0.0124 |
| | (0.027) | (0.026) | (0.020) | (0.021) |
| Treated × Duration Eligibility × Female Representation | -0.0000 | -0.0036 | 0.0066 | 0.0149 |
| | (0.043) | (0.047) | (0.027) | (0.030) |
| Parish FE | ✓ | ✓ | ✓ | ✓ |
| QOB×YOB FE | ✓ | ✓ | ✓ | ✓ |
| SES Effects | ✓ | ✓ | ✓ | ✓ |
| School Reforms | ✓ | ✓ | ✓ | ✓ |
| Parish Trends | | ✓ | | ✓ |

*** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the parish level. 'Female Representation' is a dummy variable set equal to one if the local council had more than the median share of 7 per cent females after the 1946 election. Municipalities without a council are coded as having zero female representatives. See notes of Table 3 for a list of included covariates, fixed effects and trends.

### 5.2.2 Women's Labour Supply – Childcare Expansion

We have demonstrated that growing demand for women, including as workers in the health and education sectors, contributed to realizing impacts of the infant health intervention on women's labour market outcomes. We also investigated a particular determinant of labour supply – expansion of childcare – leveraging a major expansion of state-subsidised childcare that occurred from 1963. The exposed cohorts were 30-32 years old, and this may facilitated their return to work as mothers. Details of data and method are in Appendix G.3. First, we demonstrate that childcare expansion at the local level was not a function of the infant intervention (Appendix Table G5). Second, we show that there is no evidence that growth in employment and earnings among women exposed to the infant intervention was greater in places where childcare expansion was greater (Appendix Table G6).[47]

[47]This may be because the expansion only occurred when the women were in their early 30's or it may be, as found in Norway, that the expansion only substituted out informal care (Havnes and Mogstad, 2011).

# 6 Mediators – Tracing and Decomposing Effects

Identifying mediators is a central challenge in longitudinal studies of early life interventions (Heckman et al., 2013), requiring either two sources of exogenous variation or strong assumptions regarding the relationship between treatment, mediators and main outcomes. For this reason, it has been customary to report the effects of an intervention on potential mediators alongside effects on the final outcome of interest, without attempting to weight the contributions of alternative mediators. We provided results using this approach in the preceding section.

In recent years, a number of approaches requiring less restrictive assumptions have been suggested. Identification is typically based on a sequential ignorability condition, which states that the unobserved variables that confound the relationship between the treatment and the mediator are different from those that confound the relationship between the mediator and the outcome, conditional on treatment (cf. Heckman and Pinto, 2015; Huber et al., 2017; Dippel et al., 2017). This independence assumption may be plausible in many settings, but in our case, where most outcomes considered are proxies of human capital, it seems difficult to defend such an assumption. We therefore develop a simple approach that can gauge the *relatedness* of the treatment effect of the intervention over different domains. In essence, we examine whether it is the same sub-populations that contribute to the treatment effects in different domains.[48]

We complement this analysis with the approach developed by Gelbach (2016), which leverages the omitted variable bias formula to attribute treatment effects across potential mediators. The Gelbach approach does not have the ambition of estimating causal effects, and is essentially agnostic about the causal and temporal ordering of potential mediators. Thus, if the treatment effects on different mediators are strongly correlated, the method may deliver misleading results.[49] We attempt to (partially) address this pitfall by using insights from our analysis of

[48] A similar approach has been used in Deuchert et al. (2016), but their approach requires observing the value of the mediator for treated individuals before treatment, and identification is based on this mediator having no effect on the outcome in the pre-treatment period. Thus, their approach cannot be applied to our research design.

[49] For example, if one potential mediator (e.g. high-ranking occupation) is a direct consequence of a mediator that was operative at an earlier stage of the life course (e.g. secondary schooling completion) but more strongly correlated with the main outcome (e.g. earnings), then the Gelbach approach may attribute the treatment effect to the later rather than the earlier life course variable.

correlated effects to formulate a specification for the Gelbach (2016) decomposition.

## 6.1 Attribution of Effects

In Appendix I, we show that the estimated average treatment effect on an interaction between two binary outcomes (i.e. $Y = W \cdot Z$), denoted $\tau_Y$, carries information on how strongly the treatment effects within domains defined by the two binary outcomes W and Z relate. First, we may compare $\tau_Y$ to the benchmark value $\tau_Y^{uc}$ that it would take on if the treatment effects in the two domains were completely unrelated at the individual level:

$$\tau_Y^{uc} = \tau_W \tau_Z + \tau_W \Pr\left(Z^0 = 1\right) + \tau_Z \Pr\left(W^0 = 1\right), \tag{5}$$

where $\tau_W$ and $\tau_Z$ are the average treatment effects on the two outcomes $W$ and $Z$ and $\Pr\left(Z^0 = 1\right)$ is the (estimable) counterfactual probability of observing $Z = 0$ in the treatment group in the absence of treatment, and $\Pr\left(W^0 = 1\right)$ is analogously defined.

Table 9 shows the relatedness of the treatment effects of the intervention for a number of outcomes that exhibit significant results for women. The first two columns present the estimated treatment effect on the two outcomes mentioned in the leftmost column. For example, the first row shows that exposure to the intervention is associated with an increase in the probability of scoring a high GPA in primary school (grade 4 top 20%) of 10.55 percentage points, and an increase in the probability of secondary schooling of 5.2 percentage points.[50] The third column presents $\tau_Y^{uc}$ which is the benchmark value of $\tau_Y$, the effect on the interacted outcome (top GPA *and* secondary schooling), which would be obtained if the treatment effects were uncorrelated. In this particular example, this is 3.4 percentage points. However, the unrestricted treatment effect for this joint outcome, presented in column (4), is almost twice that number, indicating that the treatment effects on the two outcomes are strongly correlated. The three rightmost

[50]These are essentially the results from the previous section. They are slightly different because slightly different samples are occasioned now by the requirement that both outcomes are observed for a given individual.

columns present the estimated correlation coefficient between the two treatment effects.[51]

Table 9 exhibits some striking patterns. First, the estimated value of $\tau_Y$ is always well above the benchmark value $\tau_Y^{uc}$, typically twice as large, suggesting that the treatment effects are strongly correlated for all pairs of outcomes (the correlation coefficient is always greater than 0.5 for the maintained assumption on compliers). Treatment effects on earnings are highly correlated with treatment effects on each of high-ranking occupation (0.60), secondary schooling (0.58) and top GPA (0.54).[52]

The preceding results suggest that a plausible sequence of events leads from better primary school performance to secondary school completion and hence better occupations and higher earnings. However, the intervention had larger impacts on primary school scores, occupation and earnings than it did on secondary school completion.[53] This suggests there may be an alternative sequence leading directly from test scores to higher earnings, independently of secondary schooling. So as to discriminate between the two paths, in the next section we estimate a relatively flexible specification that introduces interactions with secondary schooling.

## 6.2 Gelbach Mediation Analysis

We used the Gelbach (2016) approach to estimate the relative contribution of endogenous outcomes at different stages of the lifecourse to earnings in adulthood. Denoting by $Y$ a $N \times 1$ vector representing top earnings and by $T$ the $N \times 1$ a vector of treatment assignment, we may

[51]The baseline estimate is based on the assumption that individuals who are compliers for only one of the outcomes are proportionately drawn from the populations of never-takers and always-takers in the other variable. As a sensitivity check we present estimates in square brackets that are obtained with variations in this assumption, allowing that the compliers for one outcome who are never-takers for the other outcome are either strongly under-represented or strongly over-represented; see Appendix I for details.

[52]The strongest correlation in effects is found between top GPA in primary school and high-ranking occupations (defined as managers and professionals, and accounting and administration), the correlation coefficients are greater than 0.98 and robust to different assumptions regarding the distribution of compliers in the population. The next largest correlation between treatment effects for top GPA and secondary schooling, at 0.77.

[53]We showed earlier that the intervention had the following impacts: probability of a top GPA increases by about 10 percentage points and the probability of earning a top income increases by 7-8 percentage points, but secondary school completion increases by only 4 percentage points.

**TABLE 9.** CORRELATED TREATMENT EFFECTS: WOMEN

| | | (1) | (2) | (3) | (4) | (5) | | (6) | (7) |
|---|---|---|---|---|---|---|---|---|---|
| Outcome 1 | Outcome 2 | $\tau_1$ | $\tau_2$ | $\tau_Y^{uc}$ | $\tau_Y$ | | | $\text{corr}\,(\tau_{1i}, \tau_{2i})$ | |
| Top GPA | | | | | | | | | |
| | Secondary | 0.1055* | 0.0519* | 0.0337 | 0.0664*** | | | 0.7738 | |
| | | (0.062) | (0.027) | | (0.024) | [ | 0.5332 | – | 0.8426 ] |
| | High Occ | 0.1044* | 0.0631 | 0.0485 | 0.0856** | | | 0.9848 | |
| | | (0.063) | (0.056) | | (0.039) | [ | 0.9524 | – | 0.996 ] |
| | Top Income | 0.1044* | 0.0837* | 0.0465 | 0.0704* | | | 0.5420 | |
| | | (0.063) | (0.050) | | (0.041) | [ | 0.1697 | – | 0.6484 ] |
| Secondary | | | | | | | | | |
| | High Occ | 0.0396** | 0.0815** | 0.0276 | 0.0458*** | | | 0.6121 | |
| | | (0.017) | (0.038) | | (0.014) | [ | 0.2426 | – | 0.7177 ] |
| | Top Income | 0.0396** | 0.0649** | 0.0212 | 0.0392*** | | | 0.5825 | |
| | | (0.017) | (0.033) | | (0.013) | [ | 0.2213 | – | 0.6857 ] |
| High Occ | | | | | | | | | |
| | Top Income | 0.0817** | 0.0650** | 0.0376 | 0.0568** | | | 0.6005 | |
| | | (0.038) | (0.033) | | (0.024) | [ | 0.2748 | – | 0.6936 ] |

$\tau_Y^{uc}$: benchmark value, uncorrelated effects (see Appendix I for a derivation); $\tau_1$: treatment effect outcome 1; $\tau_2$: treatment effect outcome 2; $\tau_Y$: joint treatment effect for interacted outcome $1 \times 2$; $\text{corr}\,(\tau_{1i}, \tau_{2i})$: Correlation coefficient between treatment effects. Bounds for alternative assumptions in square brackets (see Appendix I for a derivation).

compare results from two specifications; one where all potential mediators $Z$ are included as covariates, and a base specification which only includes the base covariates and fixed effects $X$:

$$Y = T\tau + X\lambda + \epsilon \tag{6}$$

$$Y = T\tau + Z\beta + X\lambda + \upsilon \tag{7}$$

Let $\hat{\tau}_{base}$ denote the estimate of $\tau$ based on specification (6), and $\hat{\tau}_{full}$ denote the estimate of $\tau$ based on specification (7). As shown by Gelbach (2016), their difference $\hat{\delta} = \hat{\tau}_{base} - \hat{\tau}_{full}$

represents an estimate of how much of the estimated effect can be attributed to the mediating variables $Z$. This decomposition of the effect does not have a causal interpretation since the exogeneity assumption $\mathbb{E}(v \mid T, Z, X) = 0$ may be violated even if the base specification (6) is identified. Nevertheless, the decomposition gives an indication of the quantitative importance of potential mediators and their respective contributions to the overall treatment effect $\tau_{base}$. The contribution of variable $k$ can be quantified as $\hat{\delta}_k = \hat{\Gamma}_k \hat{\beta}_k$; where $\hat{\Gamma}_k$ represents the effect of the intervention on mediator $k$, and $\hat{\beta}_k$ is the estimate for this variable in specification (7).

As mediators $Z$, we consider all trajectories through previous stages which may precede the outcome $Y$. Thus, when the outcome considered is earning a top income in 1970, we consider each possible trajectory going through primary school performance (Top GPA yes/no), secondary schooling enrollment (yes/no) and working in a high-ranking occupation (yes/no). There are thus 8 possible trajectories, and we use the combination $N, N, N$ (no top GPA, no secondary schooling, no high-ranking occupation) as the reference category.[54]

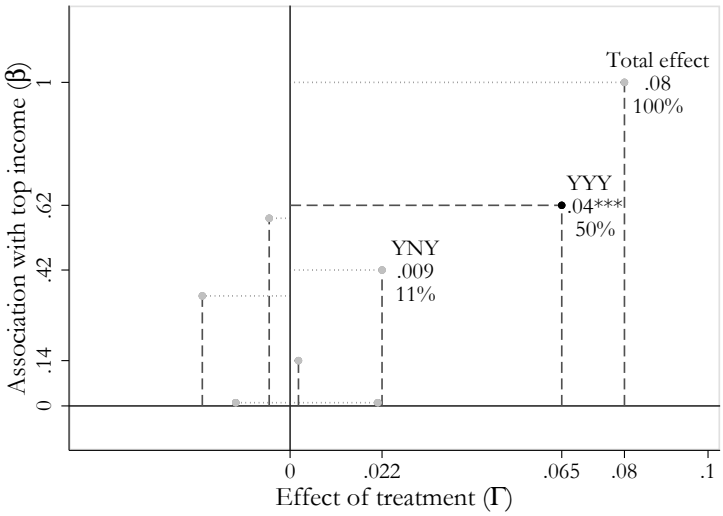
The outcome of this analysis for top income is visualised in Figure 8 (for regression results see Appendix Table J1).[55] The $X$ axis shows estimates of $\Gamma_k$s, i.e. by how much the intervention increased the probability of observing a certain outcome/mediator. The $Y$ axis shows estimates of the $\beta$s, which capture the association between the specific outcome/mediator and the primary endpoint (top income in this case). Thus, the area resulting from an interaction of these two estimates will be the estimate of the contribution of that mediator to the overall treatment effect, $\hat{\delta}_k = \hat{\Gamma}_k \hat{\beta}_k$. As reported above, the effect of the intervention on the probability of earning an income in the top quintile is $0.08$, which we denote "Total effect" in the figure. Our estimates suggest that the trajectory $YYY$ – signifying top GPA, secondary enrollment and high-ranking occupation – is responsible for half of this effect, or 4 percentage points. This is due to the intervention increasing the probability of entering this trajectory by 6.5 percentage points, and this trajectory being associated with a probability of being in the top

[54] By construction, the mediator pathways are correlated: if you follow path a, you do not follow path b. However, this does not bias the estimates since the paths are exhaustive and mutually exclusive.

[55] We use top income instead of log earnings as the main outcome here, because the estimates for $\beta$ are bounded between 0 and 1 and therefore more convenient to present graphically. Appendix Table J2 shows the corresponding results when log earnings are used as the main outcome.

income quintile of 62 percentage points. Second place is taken by the trajectory $YNY$ – which signifies having a top GPA and a high-ranking occupation, but no secondary schooling. This trajectory is associated with an increase of 42 percentage points in the probability of earning a top income, and the intervention increases the probability of entering this trajectory by 2.2 percentage points (though not statistically significant). The contribution by this trajectory is thus 0.9 percentage points or 11 per cent of the total effect. None of the other trajectories make quantitatively meaningful contributions to the total effect.

**FIGURE 8.** GELBACH MEDIATION, WOMEN: TOP INCOME




The figure shows how different trajectories defined by a) top GPA (Y/N), b) secondary schooling enrollment (Y/N) and c) high-ranking occupation (Y/N) contribute to the estimate of the overall effect of the intervention on top income. The $X$ axis measures the effect of the intervention on the mediator/outcome and the $Y$ axis measures the association between the mediator and the main outcome (top income). Dashed lines represent significance at the 1 per cent level, dash-dotted lines represent significance at the 10 per cent level, and dotted lines represent insignificant estimates. See Appendix Table J1 for regression results.

Appendix Figure J1 presents the corresponding results for the two intermediate outcomes secondary schooling and high-ranking occupation. Considering only top GPA as a determinant of secondary schooling, 69 per cent of the overall effect on secondary schooling may be attributed to this indicator or primary school performance. Considering high-ranking occupations, it is the trajectory combining top GPA and secondary schooling that explains most (39 per cent) of the overall effect. Second place is the trajectory $NY$ – secondary schooling graduates who did not have a top GPA. Corresponding results for males are available in Appendix

Tables J3 and J4.

# 7 Robustness Checks

**Treatment indicator.** We investigate alternative (binary) treatment indicators. The pattern of results is in general robust to these variations, and we learn that early exposure (at 0-3 months) is most effective in modifying outcomes (Appendix Table K1 and K2).

**Dropping covariates**. Our preferred specification includes several control variables. Appendix Table K3 presents estimates when excluding those covariates. Reassuringly the results are very similar to those of our main specification.

**Alternative clustering**. Our choice of clustering at the parish level was guided by the data collection process: the matching algorithm was run at the parish level because this is the administrative level for which census data were available. The consequence is that whereas there is one control parish (or, in urban areas, city) for each treated parish (or city), there are 38 different health districts in the control group. It would thus be less straightforward to conduct randomisation inference or bootstrapping at this level, and covariate balance could not be imposed. However, since the treatment exposure varies at the health district level we also considered clustering at this level. Results are provided in Appendix Table K4 and are very similar to those of our main table.

**Dropping cities**. There were only two cities included among the treated districts. In order to safeguard against results being driven by the control cities representing poor matches, we present estimates based on only rural parishes in Appendix Table K5. Dropping the cities, which represent 20 per cent of the sample, does not affect the results.

**Anchoring of grading scale**. We investigated sensitivity of the results for academic achievement to the grading scale by anchoring the scale to log income (Bond and Lang, 2013; Cunha and Heckman, 2008). This gives mean log income for each of the seven steps on the scale. Results are in Appendix Table K6. The correlations in fourth grade imply that a switch of test scores from 1 to 6 points associates with an earnings gain of 95%. Appendix Table K7 shows regression results using income. The estimates are similar to those using the grade scale.

Anchoring with years of education generates similar results.

**Uncertain matches**. The dataset on school performance was linked to the other sources based on birth parish, names, and exact date of birth. Individuals moving away from their birth parish were tracked in church books and then added manually to the dataset. Our record linkage algorithm allows for minor deviations in dates or spelling of names; however, when the birth parish is different from the school parish, there is a greater risk of false positives and therefore we were more restrictive in this part. As a robustness check, we add 338 uncertain matches and re-estimate the effects on school performance. Appendix Table K8 shows that results are generally insensitive to this extension of the dataset.

**Balancing tests**. Panel A Appendix Table K9 presents regression results from a specification where we replace the outcome variables with baseline characteristics – mother's age and marital status, household head occupational category, whether the child has a younger sibling in the dataset, whether born in hospital, and the compulsory schooling rules applying to their cohort in the birth parish. In addition, Panel B shows the result for a couple of aggregate variables, such as the birth rate in the birth parish, and the share of midwife-assisted births in their health district in their birth year. Only one of the ten variables is significantly correlated with eligibility – the mother's marital status – but the coefficient is very small. Nevertheless, this result justifies the inclusion of marital status among the control variables.

**Selective survival.** To investigate the role of survival selection, we estimate intervention effects on earnings for sub-samples of individuals surviving until age 40, 50, 60, 70 and 75 (Appendix Table K10). Our main findings are not significantly different across the samples.

**Pre-trends test.** To investigate whether the outcomes of interest followed similar trends in the treatment and control regions before the intervention, we use the pre-intervention sample to estimate the following equation:

$$y = \beta(trend \times treated) + \gamma treated + \delta trend + \varepsilon.$$

$trend$ is a trend variable based on each $month \times year$ observation in the pre-intervention sample and $treated$ is an indicator for treated parishes. A premise of our strategy is that $\beta$ equals zero. Appendix Tables K11 and K12 show results for primary school and labour market

outcomes. In general we cannot reject that $\beta$ equals zero. The fact that our results are robust to parish specific time trends also suggests the absence of differential pre-trends.

**Attrition**. Match rates are reported in Section 3. Match rates of the birth sample with the 1970 census and the tax register are high, but the dominant cause of attrition is death. Since we know from Bhalotra et al. (2017) that the intervention of interest altered child and adult mortality risks, there is a potential concern about differential attrition. The match rate with the school data is lower since, in addition to death, some of the archives did not preserve all school catalogues. Appendix Table K13 provides an overview of attrition rates for sub-samples identified by treated parish and eligible birth date. Table 10 provides tests for whether attrition is systematically related to treatment. Attrition is negatively associated with treatment in the school sample, but uncorrelated with treatment in the later-life samples.[56]

**TABLE 10.** ATTRITION IN DIFFERENT SAMPLES

| | School Sample | | | Census 70 Sample | | | Pension Sample | | |
|---|---|---|---|---|---|---|---|---|---|
| Treated Parish | 0.0450 | | | 0.0014 | | | -0.0167 | | |
| | (0.047) | | | (0.012) | | | (0.015) | | |
| Duration Eligibility | 0.0200 | 0.0985 | 0.1048 | 0.0193** | 0.0133 | 0.0142 | 0.0146 | 0.0553 | 0.0458 |
| | (0.015) | (0.077) | (0.076) | (0.008) | (0.074) | (0.073) | (0.021) | (0.087) | (0.087) |
| Treated × Duration Eligibility | -0.0387* | -0.0370* | -0.0562** | -0.0022 | -0.0025 | -0.0100 | 0.0007 | -0.0014 | -0.0101 |
| | (0.021) | (0.020) | (0.023) | (0.009) | (0.009) | (0.009) | (0.023) | (0.021) | (0.020) |
| Baseline | 0.339 | 0.339 | 0.339 | 0.140 | 0.140 | 0.140 | 0.336 | 0.336 | 0.336 |
| Parish FE | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| QOB×YOB FE | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| School FE | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| SES Effects | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Length of Schoolyear | | ✓ | ✓ | | | | | | |
| Schoolform | | ✓ | ✓ | | | | | | |
| Parish Trends | | | ✓ | | | ✓ | | | ✓ |

*** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the parish level. See notes of Table 2 for definitions of fixed effects and trends.

[56]Wüst et al. (2018) find that the Danish home visiting program increased the probability of emigration. The result that attrition is uncorrelated with treatment in our samples using administrative register data on later life outcome, suggest the infant care intervention did not have such an impact on emigration out of Sweden.

We address this in two ways. First, we ran the analysis of census outcomes on a sample restricted to individuals who also appear in the school sample (Appendix Table K14). We see positive and significant effects of the intervention on women's education and earnings of magnitudes similar to the baseline estimates.[57] Second, we estimated Lee bounds for all main outcomes (Appendix Tables K15, K16). The results stand up to accounting for attrition in this way. With the one exception of Top GPA, for which the lower Lee bound has a $p$ value of 0.11, all of the positive effects reported for women are significant even at the lower Lee bound.

**Placebos**. For the 1970 census outcomes we implement a placebo test, using a fake intervention ten years after the actual intervention. For this, we generated a sample of children born in treatment and control areas ten years after the infant intervention, in 1940-44, using the 1950 population census and the Swedish Death Index. The only covariate available from these sources is the individual's sex. Generating a fake treatment group using parish and date of birth, we show that the fake intervention had no long run impacts (Appendix Table K17).[58]

**Randomisation inference**. We conduct a randomisation inference test for long-term outcomes, also in the spirit of a placebo test. We randomly assign treatment status within each treatment and control parish pair using 5,000 permutations; cf. Karlsson and Pichler, 2015 for a discussion of randomisation inference in DiD settings. Following MacKinnon and Webb (2016) we present results based on $t$ statistics. Appendix Figures K1 and K2 plot the distributions of placebo treatment effects and display the actual treatment effect and the corresponding $p$ value. Except for part-time employment for women, where the distributions do not look smooth, the results are similar to the main estimates in Tables 3, 4 and 5.

[57]The exercise in Section 6 was also on this restricted sample, where we saw the results hold.

[58]In order to use the years 1934-1939, we would need to digitize additional years of parish birth records and school exam catalogues. The results for the 1940–1944 cohorts have to be viewed with some caution since before 1947 the parish of birth that was reported refers to the location of the hospital the birth was in and not to the place of registration of the parents (Holmlund, 2008). With a rising share of institutionalised births over time this leads to some misreporting for our placebo test cohorts. We do not face this problem for cohorts born 1930–1934 since the parish records that were digitised within this project report the place of registration of the parents and not the place of the hospital they were born in. To mitigate the problem, we control for hospital births.

# 8 Discussion – Related Literature

In the Introduction, we highlighted our main contribution. Here we discuss a related literature and delineate our contributions to specific domains. We contribute to evidence demonstrating that early life health interventions can have a causal impact on cognitive attainment. Figlio et al. (2014) state *"While we have strong evidence from twin comparison studies that poor initial health conveys a disadvantage in adulthood, we have little information about the potential roles for policy interventions in ameliorating this disadvantage during childhood"*. Figlio et al. (2014) and Black et al. (2007) use sibling or twin estimators to identify impacts of birth weight on cognitive performance in Norway and Florida respectively. Like Figlio et al. (2014), we are able to assess impacts of infant health on cognitive scores at different ages and by the socio-economic characteristics of parents. However, while they analyse impacts of birth weight differences, we use population-level exposure to an intervention that improved infant health.[59] Our evidence thus makes a case for the many emerging policies in rich and poor countries that target the health of the newborn child. Other studies that show cognitive gains from infant health policies include Chay et al. (2009) who study black-white convergence in test scores as a function of hospital de-segregation in America, Bharadwaj et al. (2013) who show impacts of neonatal care facilities on school test scores in Chile and Norway, Bhalotra and Venkataramani (2013) who demonstrate impacts of infant exposure to a clean water programme in Mexico on cognitive attainment in middle and late adolescence

Our findings also contribute to a scarce literature providing evidence that cognitive performance and higher education contribute to earnings. Pre-school programmes such as Project STAR and the Perry intervention appear to have raised long term earnings by generating improvements not in cognitive skills but, instead, in health and non-cognitive skills (Chetty et al., 2011; Heckman et al., 2013, 2006; Baker et al., 2018). A vast body of research documents long run benefits of early life health interventions on earnings (Almond and Currie, 2011; Heckman et al., 2014; Falk and Kosse, 2016; Bütikofer et al., 2019). While it is implicit that the intervening mechanism is human capital accumulation, there is fairly limited evidence of the

[59] Our data contain records of sickness-related absence from school, allowing us to argue that the mechanism was linked to early life health rather than contemporaneous health.

importance of cognitive skills in this process. A reason is that few previous studies have been able to link data on test scores and educational choices to adult earnings. The seminal study in this domain is Black et al. (2007) who use twin-comparisons to show that IQ and earnings in adulthood are both increasing in birth weight. They do not have any intermediate outcome data.

A third contribution is to a literature showing gender differences in dynamic responses to early life interventions (Baird et al. (2016); Bhalotra and Venkataramani (2013); Garcia et al. (2018); Molina (2020); Bobonis et al. (2006); Maluccio et al. (2009); Maccini and Yang (2009); Field et al. (2009)). We show how differences in skill accumulation and opportunities mattered. Greater skill accumulation among girls is consistent with the theoretical framework of Pitt et al. (2012), premised on men having a comparative advantage in brawn-intensive tasks and women in tasks that are relatively intensive in cognitive function. We showed that girls in our study cohorts faced higher returns to secondary schooling. Additional reasons are that the cognitive growth curve differs by gender, that girls have stronger non-cognitive skills (like conscientiousness) so that for a given increment to the cognitive endowment, they work harder.[60]

Our work relates to recent studies examining long-run impacts of similar mother-baby programmes in Denmark and Norway (Hjort et al., 2017; Wüst et al., 2018; Bütikofer et al., 2019). Hjort et al. (2017) and Wüst et al. (2018) differ from this paper because they look only at impacts on adult health, similar to Bhalotra et al. (2017), who show that exposure to the infant intervention led to reduced adult deaths from infections, cancer and cardiovascular disease. Bütikofer et al. (2019) is more directly related to this paper as it studies years of education and labour market outcomes. We broadly reinforce their finding for Norway that an infant health programme led to higher earnings on average. However, there are several differences in context

[60]In a recent review Almond et al. (2017) argue that the effects of the early life environment on long run outcomes are often heterogeneous, *"reflecting differences in child endowments, budget constraints, and production technologies"*. Here, we additionally highlight the potential role of opportunities, which are context-dependent. In a similar spirit, Coles and Francesconi (2017) argue that an expansion of job opportunities for women was critical to realisation of the impacts of the pill innovation on women's labour market outcomes in America, and Bhalotra and Venkataramani (2012) show that labour market segregation in the Southern states of America limited realisation of earnings gains from infant exposure to antibiotics for black but not white Americans.

and approach relative to Bütikofer et al. (2019). In particular, the following contributions are all relevant to our being able to significantly advance the analysis of mechanisms. First, we have school test score data. Second, we document not just impacts on earnings but also on employment, sector and occupation. Third, we estimate treatment effects across the distribution of test scores and income. Having identified systematic differences in higher education and labour market performance between men and women, we supplement a descriptive mediation exercise with well-identified mechanism analyses showing (i) how the treatment led to boys being crowded-out of secondary school places, and (ii) how predicted demand moderated labour market gains for women relative to men. The story we tell is thus very different.

Methodological differences include that Bütikofer et al. (2019), Hjort et al. (2017) and Wüst et al. (2018) analyse nationwide infant care programmes rolled out over decades, while ours was a short pioneering trial announced as concluding within two years, and this limited the possible influence of confounders and unobserved trends (cf. Goodman-Bacon, 2018; de Chaisemartin and d'Haultfoeuille, 2019). Also, using rich information on parental characteristics digitised from church records, we can show that family socioeconomic status is balanced between treated and non-treated individuals.[61]

# 9 Conclusion

Using unique longitudinal data in which individual outcomes are observed at different stages of the life course, we identify large impacts of a universal infant care intervention on school and labour market outcomes. A crude estimate of the internal rate of return suggests that the trial was highly cost-effective, and it was successfully scaled up following the short trial period that we analyse. Our findings are of contemporary relevance given that poor health and nutrition and deficient early childhood care are estimated to be causing about 200 million children under the age of 5 to fail to attain their cognitive potential, and that this has been identified as a key

[61] Most studies examining the long-run effects of early-life shocks do not have the data to check that treated individuals are statistically exchangeable with non-treated individuals with respect to family SES. Brown and Thomas (2018) show that if parental SES varies with treatment, then not adjusting for it may generate spurious results.

factor in the intergenerational transmission of poverty (Grantham-McGregor et al., 2007).

Intervention effects are highly correlated across outcomes, implying that it is largely the same individuals who drive the various effects. With the caveat that it is only descriptive, our analysis of mediators suggests that cognitive attainment, secondary schooling and having a high-ranking occupation contributed significantly to the increase in adult earnings. The analysis also highlights the importance of (a) looking at the distribution of test score and earnings outcomes and (b) the relevance of institutional capacity and demand conditions. It shows that population health improvements can lead to a demand for higher education, and that the impact of infant health interventions on earnings will tend to be larger when the demand for skilled workers is rising. Our results also highlight that gender segmentation in the labour market, which may lead to differential demand growth for male and female labour, may contribute to understanding the tendency for infant health interventions to have different impacts on the earnings of men and women.

In our setting, as will be the case in many developing countries today, secondary school places were limited and entry was competitively determined. Although intervention effects on boy test scores were stronger on average, it mattered that treatment effects on the chances of scoring in the top quintile were significantly greater for girls. This led to intervention effects on girls progressing into secondary schooling with no corresponding effect for boys. Our analysis suggests that treated boys were thereafter on a lower trajectory than treated girls. Leveraging variation in the share of treated girls in secondary school catchment areas we demonstrate displacement of boys. In addition, we show that when these cohorts emerged onto the labour market, the movement of high-skilled women into high-earning sectors was facilitated by growth in labour demand for skilled women, a large part of which was in woman-friendly public sector jobs created by expansion of the welfare state.

# References


Acemoglu, D., D. H. Autor, and D. Lyle (2004). Women, war, and wages: The effect of female labor supply on the wage structure at midcentury. *Journal of political Economy 112*(3), 497–551.

Adhvaryu, A. and A. Nyshadham (2016). Endowments at birth and parents' investments in children. *The Economic Journal 126*(593), 781–820.

Aizer, A., J. Currie, P. Simon, and P. Vivier (2018). Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics 10*(1), 307–41.

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives 25*(3), 153–172.

Almond, D., J. Currie, and V. Duque (2017). Childhood circumstances and adult outcomes: Act ii. *Forthcoming Journal of Economic Literature*.

Almond, D., B. Mazumder, and R. Van Ewijk (2014). In utero ramadan exposure and children's academic performance. *The Economic Journal 125*(589), 1501–1533.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

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

Attanasio, O. P., C. Fernandez, E. Fitzsimons, S. Grantham-McGregor, C. Meghir, and M. R. Codina (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in colombia: cluster randomized controlled trial. *British Medical Journal 349*(g5785).

Aucejo, E. and T. Romano (2014). Assessing the Effect of School Days and Absences on Test Score Performance. CEP Discussion Papers 1302, Centre for Economic Performance, London School of Economics.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279–1333.

Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness 6*(1), 1–43.

Baird, S., J. H. Hicks, M. Kremer, and E. Miguel (2016). Worms at work: Long-run impacts of a child health investment. *The Quarterly Journal of Economics 131*(4), 1637–1680.

Baker, M., J. Gruber, and K. Milligan (2018). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy Forthcoming*.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics 122*(3), 1235–1264.

Bång, J. (2001). The Returns to Education - Using Data Retrieved from the Swedish National Census of 1930. *mimeo, Department of Economics, Uppsala University*.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science 335*(6068), 582–586.

Benjamini, Y., A. Krieger, and D. Yekutieli (2006). Adaptive linear step-up false discovery rate controlling procedures. *Biometrika 93*(3), 491–507.

Bernhardtz and Klintfelt (2007). Det typiska adoptivbarnet. *Välfärd 2*.

Bhalotra, S., M. Fernandez-Sierra, and A. S. Venkataramani (2018). Women's labour force participation and the gender wage gap across the distribution of tasks.

Bhalotra, S., M. Karlsson, and T. Nilsson (2017). Infant health and longevity: Evidence from a historical intervention in sweden. *Journal of the European Economic Association*, jvx028.

Bhalotra, S., M. Karlsson, T. Nilsson, and N. Schwarz (2019). Infant health, cognitive performance and earnings: Evidence from inception of the welfare state in sweden.

Bhalotra, S., A. Venkataramani, and S. Walther (2018). Fertility responses to reductions in mortality: Quasi-experimental evidence from 20th century america.

Bhalotra, S. R. and A. S. Venkataramani (2012). Shadows of the captain of the men of death: Early life health interventions, human capital investments, and institutions. *Available at SSRN*.

Bhalotra, S. R. and A. S. Venkataramani (2013). Cognitive development and infectious disease: Gender differences in investments and outcomes. *IZA Discussion Paper* (7833).

Bharadwaj, P., J. Eberhard, and C. Neilson (2017). Health at birth, parental investments and academic outcomes. *Forthcoming in Journal of Labour Economics*.

Bharadwaj, P., K. V. Løken, and C. Neilson (2013). Early life health interventions and academic achievement. *American Economic Review 103*(5), 1862–1891.

Bitler, M., T. Domina, and H. Hoynes (2016). Experimental Evidence on Distributional Effects of Head Start. Mimeo, UC Davis.

Björklund, A. and C. Kjellström (1994). Avkastningen på utbildning i sverige 1968 till 1991.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007). From the cradle to the labor market? the effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, 93–120.

Bobonis, G. J., E. Miguel, and C. Puri-Sharma (2006). Anemia and school participation. *Journal of Human resources 41*(4), 692–721.

Bond, T. N. and K. Lang (2013). The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results. *The Review of Economics and Statistics 95*(5), 1468–1479.

Borgen, N. T. (2016). Fixed effects in unconditional quantile regression. *The Stata Journal 16*(2), 403–415.

Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research.

Bozzoli, C., A. Deaton, and C. Quintana-Domeque (2009). Adult height and childhood disease. *Demography 46*(4), 647–669.

Brown, R. and D. Thomas (2018). On the long term effects of the 1918 us influenza pandemic. *Mimeo University of Colorado Denver*.

Bütikofer, A., K. V. Løken, and K. G. Salvanes (2019). Infant health care and long-term outcomes. *Review of Economics and Statistics 101*(2), 341–354.

Bygdeman, M. and K. Lindahl (1994). Sexualupplysning och reproduktiv hälsa i sverige under 1900-talet. *Sveriges Offentliga Utredningar*.

Cattan, S., G. Conti, and C. Farquharson (2019). Workforce quality in early years intervention: Evidence from a large-scale home-visiting program.

Cattan, S., D. A. Kamhöfer, M. Karlsson, and T. Nilsson (2017). The short-and long-term effects of student absence: Evidence from sweden. *IZA Discussion Paper* (10995).

Chay, K. C., J. Guryan, and B. Mazumder (2009). Birth cohort and the black-white achievement gap: The roles of access and health soon after birth. (Working paper 15078).

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

Clarke, D., G. Cortés Méndez, and D. Vergara Sepúlveda (2018). Growing together: Assessing equity and efficiency in an early-life health program in chile.

Coles, M. G. and M. Francesconi (2017). Equilibrium search and the impact of equal opportunities for women. *Journal of Political Economy*.

Cortes, G. M., N. Jaimovich, and H. E. Siu (2018). The" end of men" and rise of women in the high-skilled labor market. Technical report, National Bureau of Economic Research.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources 43*(4).

Dahr, E. (1945). Lärjungevalet till studielinjer med den nuvarande realskolans mål. *SOU 1945 44*.

Datta Gupta, N., N. Smith, and M. Verner (2006). Child Care and Parental Leave in the Nordic Countries: A Model to Aspire to? IZA Discussion Papers 2014.

De Chaisemartin, C. (2012). Fuzzy differences in differences. Pse working papers, HAL.

de Chaisemartin, C. and X. d'Haultfoeuille (2019). Two-way fixed effects estimators with heterogeneous treatment effects. Technical report, National Bureau of Economic Research.

Deuchert, E., M. Huber, and M. Schelker (2016). Direct and indirect effects based on difference-in-differences with an application to political preferences following the vietnam draft lottery.

Deverman, B. E. and P. H. Patterson (2009). Cytokines and Cns Development. *Neuron 64*, 61–78.

Dhamija, G. and S. Gitanjali (2019). Impact of early life interventions on later life outcomes: Evidence from india's integrated child development services.

Dippel, C., R. Gold, S. Heblich, and R. Pinto (2017). Instrumental variables and causal mechanisms: Unpacking the effect of trade on workers and voters. Technical report, National Bureau of Economic Research.

Doyle, O., C. P. Harmon, J. J. Heckman, and R. E. Tremblay (2009). Investing in early human development: Timing and economic efficiency. *Economics & Human Biology 7*(1), 1–6.

Duflo, E. and R. Hanna (2005). Monitoring works: Getting teachers to come to school. NBER Working Papers 11880, National Bureau of Economic Research, Inc.

Egg-Henriksen, F. (2008). As different as two drops of water? *Statistics Norway*.

Elsby, M. W., B. Hobijn, and A. Sahin (2010). The labor market in the great recession. Technical report, National Bureau of Economic Research.

Engle, P. L., M. M. Black, J. R. Behrman, M. Cabral de Mello, P. J. Gertler, L. Kapiriri, R. Martorell, and M. E. Young (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet 369*(9557), 229–242.

Eppig, C., C. L. Fincher, and R. Thornhill (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society of London B: Biological Sciences 277*(1701), 3801–3808.

Falk, A. and F. Kosse (2016). Early childhood environment, breastfeeding and the formation of preferences.

Field, E., O. Robles, and M. Torero (2009). Iodine deficiency and schooling attainment in tanzania. *American Economic Journal: Applied Economics 1*(4), 140–69.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review 104*(12), 3921–3955.

Finch, C. E. and E. M. Crimmins (2004). Inflammatory exposure and historical changes in human life-spans. *Science 305*(5691), 1736–1739.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Fischer, M., M. Karlsson, T. Nilsson, and N. Schwarz (2019). The long-term effects of long terms. compulsory schooling reforms in sweden. *Forthcoming in Journal of European Economic Association*.

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

Fredriksson, V., S. Marklund, G. Sivgard, and M. Widen (1971). Svenska folkskolans historia, sjätte delen. skolutvecklingen 1942–1962.

Garcia, J. L., J. J. Heckman, and A. L. Ziff (2018). Gender differences in the benefits of an influential early childhood program. *European economic review 109*, 9–22.

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

Goldin, C. (1988). Marriage bars: Discrimination against married women workers, 1920's to 1950's.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018). Bartik instruments: What, when, why, and how. Technical report, National Bureau of Economic Research.

Goodman, J. (2014). Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time. NBER Working Papers 20221, National Bureau of Economic Research, Inc.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research.

Gorna, R., N. Klingen, K. Senga, A. Soucat, and K. Takemi (2015). Women's, children's, and adolescents' health needs universal health coverage. *The Lancet 386*(10011), 2371–2372.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet 369*(9555), 60–70.

Havnes, T. and M. Mogstad (2011). Money for nothing? universal child care and maternal employment. *Journal of Public Economics 95*(11-12), 1455–1465.

Heckman, J., D. Ohls, R. Pinto, and M. Rosales (2014). A reanalysis of the nurse family partnership program: The memphis randomized control trial. `https://heckman.uchicago.edu/sites/heckman2013.uchicago.edu/files/uploads/CEHD_Launch/2_CEHD-NFP_SLIDES_2014-05-29a_MR_FINAL.pdf`.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–86.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science 312*(5782), 1900–1902.

Heckman, J. J. and R. Pinto (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews 34*(1-2), 6–31.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Hinnerich, B. T. and P. Pettersson-Lidbom (2014). Democracy, redistribution, and political participation: Evidence from sweden 1919–1938. *Econometrica 82*(3), 961–993.

Hjalmarsson, R., H. Holmlund, and M. J. Lindquist (2015). The effect of education on criminal convictions and incarceration: Causal evidence from micro-data. *The Economic Journal 125*(587), 1290–1326.

Hjort, J., M. Sølvsten, and M. 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.

Holmlund, H. (2008). A Researchers Guide to the Swedish Compulsory School Reform. CEE Discussion Papers 0087, Centre for the Economics of Education, LSE.

Huber, M., M. Lechner, and A. Strittmatter (2017). Direct and indirect effects of training vouchers for the unemployed. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Imbens, G. and J. Woolridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of public Economics 89*(5-6), 761–796.

Karlsson, M. and S. Pichler (2015). Demographic consequences of HIV. *Journal of Population Economics 28*(4), 1097–1135.

Killewald, A. and J. Bearak (2014). Is the motherhood penalty larger for low-wage women? a comment on quantile regression. *American Sociological Review 79*(2), 350–357.

Leeuwen, v. M., I. Maas, and A. Miles (2002). Hisco: Historical international standard classification of occupations. *Leuven University Press*.

Levine, P. B. and D. Schanzenbach (2009). The impact of children's public health insurance expansions on educational outcomes. In *Forum for Health Economics & Policy*, Volume 12. De Gruyter.

Lindgren, K.-O., S. Oskarsson, and C. T. Dawes (2014). Can political inequalities be educated away? Evidence from a Swedish school reform. Technical report, Working Paper, IFAU-Institute for Evaluation of Labour Market and Education Policy.

Lindgren, K.-O., S. Oskarsson, and M. Persson (2019). Access to education and political candidacy: Lessons from school openings in sweden. *Economics of Education Review 69*, 138–148.

Maccini, S. and D. Yang (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review 99*(3), 1006–26.

MacKinnon, J. G. and M. D. Webb (2016). Randomization Inference for Difference-in-Differences with Few Treated Clusters. Working Papers 1355, Queen's University, Department of Economics.

Maluccio, J. A., J. Hoddinott, J. R. Behrman, R. Martorell, A. R. Quisumbing, and A. D. Stein (2009). The impact of improving nutrition during early childhood on education among guatemalan adults. *The Economic Journal 119*(537), 734–763.

Molina, T. (2020, 03). Pollution, Ability, and Gender-Specific Investment Responses to Shocks. *Journal of the European Economic Association*. jvaa005.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy 119*(1), 39 – 77.

Nowakowski, R. S. (2006). Stable neuron numbers from cradle to grave. *Proceedings of the National Academy of Sciences of the United States of America 103*(33), 12219–12220.

Pettersson-Lidbom, P. (2015). Midwives and Maternal Mortality: Evidence from a Midwifery Policy Experiment in Sweden in the 19th Century.

Pitt, M. M., M. R. Rosenzweig, and M. N. Hassan (2012). Human capital investment and the gender division of labor in a brawn-based economy. *The American Economic Review 102*(7), 3531–3560.

Population and Housing Census 1970 (1972). Part 10. industry, occupation and education in the whole county etc. *National Bureau of Statistics*.

Saaritsa, S. and A. Kaihovaara (2016). Good for girls or bad for boys? schooling, social inequality and intrahousehold allocation in early twentieth century finland. *Cliometrica 10*(1), 55–98.

SCB (1972). Population and housing census 1970, part 1. population in communes and parishes, etc. *National Bureau of Statistics*.

Sievertsen, H. H. and M. Wüst (2017). Discharge on the day of birth, parental response and health and schooling outcomes. *Journal of health economics 55*, 121–138.

SOU (1929). Betänkande angående moderskapsskydd. *Statens offentliga utredningar, Stockholm 1929:28*.

SOU (1935). Kungl. medicinalstyrelsens utlåtande och förslag angående förebyggande mödra och barnavård. *Statens offentliga utredningar 1935:19*.

SOU (1942). Betänkande med utredning och förslag angående betygsättningen i folkskolan, angivet av inom ecklesiastikdepartementet tillkalade sakkunniga. *Statens offentliga utredningar 1942:11*.

Stanfors, M. (2003). *Education, labor force participation and changing fertility patterns : a study of women and socioeconomic change in twentieth century Sweden*, Volume 22. Lund studies in economic history.

Statistics Sweden (1937). Folkräkningen den 31 december 1930. *I. Areal, folkmängd och hushåll inom särskilda förvaltningsområden m.m. befolkningsagglomerationer 2*.

Statistics Sweden (1947). Kommunala valen. år 1946. Stockholm.

Stenhoff, G. (1934). Försöksverksamhet beträffande för- och eftervård vid barnsbörd. *Tidskrift för Barnavård och Ungdomsskydd 3*, 99–101.

Sundén, A. (2006). The swedish experience with pension reform. *Oxford Review of Economic Policy 22*(1), 133–148.

Swedish Government, S. (1931). Government proposition 1931:306. *Stockholm*.

UNESCO (2014). *Teaching and learning: Achieving quality for all*, Volume 2013/4. UNESCO Global Monitoring Report.

Wallgren, A. (1936). Den förebyggande spädbarnsvårdens organisation. *Tidskrift för Barnavård och Ungdomsskydd* (5), 163–67.

Wallin, H. and H. Grimlund (1933). års förnyade läroverksstadga: med förklaringar och hänvisningar: jämte timplaner och undervisningsplan mm rörande allmänna läroverken.

World Bank (2015). *Mind, Society and Behaviour. World Development Report 2015*. World Bank Group.

Wüst, M., E. L. Mortensen, M. Osler, and T. I. Sørensen (2018). Universal infant health interventions and young adult outcomes. *Health economics 27*(8), 1319–1324.

Yi, J., J. J. Heckman, J. Zhang, and G. Conti (2015). Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal 125*(588), F347–F371.