Early Life Health Interventions:

Effects on Sickness Absence and Academic Performance

Sonia Bhalotra
University of Essex**

Martin Karlsson
University of Duisburg-Essen*
University of Oslo

Therese Nilsson
Lund University†
Research Institute of Industrial Economics (IFN)

Nina Schwarz
University of Duisburg-Essen‡

April 29, 2015
preliminary version!

*University of Duisburg-Essen, Schützenbahn 70, 45127 Essen, Germany, e-mail: martin.karlsson@uni-due.de

†Department of Economics, Lund University, Box 7082, SE-220 07 Lund, Sweden, and Research Institute of Industrial Economics (IFN), Box 55665, SE-102 15 Stockholm, Sweden e-mail: therese.nilsson@nek.lu.se

**Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, e-mail: srbhal@essex.ac.uk

‡Corresponding author; University of Duisburg-Essen, Schützenbahn 70, 45127 Essen, Germany, e-mail: nina.schwarz@uni-due.de

Acknowledgements: The authors would like to thank Riksbankens Jubileumsfond (The Swedish Foundation for Humanities and Social Sciences) P12-0480:1, and the Center for Economic Demography, Lund University for generous support, and Johanna Ringkvist and Ines Hußmann for excellent research assistance.

Abstract

This paper evaluates the effects of an early life health intervention on health and academic performance among children of primary school age. We use individual level data from parish records for Sweden in the 1930s, when a field trial on maternal and infant care was rolled out in seven medical districts. We match birth records and school records on academic performance and sickness absence for pupils in grade one and grade four and validate our results by linking the applied grading scale with adult outcomes. Our estimates suggest that the infant intervention had significant effects on academic attainment for boys in grade four in 'writing' and 'reading and speaking'. We find no significant effects on sickness absence.

Keywords: Maternal care; Infant care; Education; Early life interventions; Programme evaluation; Sweden.

JEL classification: I15; I18; H41

1 Introduction

This paper evaluates the effects of early life health interventions on health and academic performance in primary school. Using individual level data from Sweden in the 1930s, we study how a field trial on maternal and infant care affected academic performance and sickness absence for pupils in grade one and grade four.

It is generally accepted that health and educational performance are correlated but the question of causation is still a matter of investigation. Whereas other studies try to evaluate whether education might causally affect health (Grossman, 1972) or whether there are third factors (e.g. time preferences) which determine both health and education (Fuchs, 1982), we will focus on the third possible link: whether health causally affects education. Exploiting the field trial as a source of exogenous variation in health, we analyse whether health has a causal effect on primary education. Furthermore, we have the possibility to investigate whether improved health conditions in the early phase of life have an effect on later life health outcomes.

There is already a vast literature on the field of how child health affects educational attainment but several of these studies have limitations like small samples or outcome variables, like completed years of schooling, which lack variation. Some of these papers exploit early life health shocks like the Spanish Flu of 1918 (Almond, 2006) or other natural experiments which expose cohorts prenatally to radiation or pollution (Black et al., 2013; Almond et al., 2009; Sanders, 2012). Other contributions look at how birth weight (Behrman and Rosenzweig, 2004; Royer, 2009; Black et al., 2007) or breastfeeding and nutrition (Del Bono and Rabe, 2012; Glewwe et al., 2001) influence later academic achievement. Most of these studies find significant effects on their educational measurements which include e.g. high school completion, completed years of education or test scores. Bharadwaj et al. (2013) for example analyse whether advanced medical technology in neonatal and childhood care in the treatment of very low birth weight children in Chile and Norway causally affects test scores. Their dataset covers the period 2002-2010 for Chile and 1986-2008 for Norway. They find significant improvements in test scores and marks at school in the magnitude of 0.15-0.22 standard deviations.

We are able to contribute to the existing literature in various ways. First, we examine positive health interventions, but unlike Bharadwaj et al. (2013), who focus on modern postnatal care, we look at basic services as currently applied in many low income as well as developed countries. Since in developing countries the consequences of early life circumstances tend to have larger long-term effects, the improvement in basic infant care is of special importance there (cf. Currie and Vogl, 2013). The state of public health, indicated by infant and maternal mortality, in current developing countries is broadly similar to that in Sweden in the early twentieth century (Razavi, 2012).

A second advantage of our study is that we observe a wide sample of treated children and not just children at the low birth weight margin. Together with the almost universal enrollment of individuals in the treated regions (about 2,000 mothers and 2,600 children enroled), this enables representative and informative conclusions, possibly relevant for policy-makers.

Third, our unique and purpose built dataset contains a large number of observations and rich information which allows us to go beyond the basic analysis of early life circumstances. Early life health can have long-lasting effects over the entire life cycle (Almond and Currie, 2011; Currie and Rossin-Sater, 2014) and early life years represent a crucial period for the development of cognitive and non-cognitive skills (Cunha and Heckman, 2007). Yet, not very much is known on how cognitive development at *different ages* might depend on early life health (Figlio et al., 2014). Our purpose built dataset allows us to investigate this continuity in academic performance for pupils in their first and fourth school year when they are between 7 and 11 years old.

Fourth, beside of marks in grade one and four, the data also allows us to study whether the early life health intervention had an effect on sickness absence at the same time. Therefore we have the opportunity to explore the interdependence between sickness absence and marks at a later stage of evaluation.

Fifth, due to the possibility of a long term follow up of the individuals and the comprehensiveness of our dataset, we are also able to anchor the applied grading scale to later life earnings (labour income in 2002) which validates the findings. Since marks are given on an ordinal scale it is necessary to control whether the effects are only due to the choice of the scale (Bond and Lang, 2013).

The field trial involved seven medical districts in Sweden which consisted of 57 parishes and 2 cities. We therefore gathered data from individual birth records of these seven treated health districts and corresponding birth record data from matched control districts in Sweden for the years 1930-1934. We also collect and digitise school performance data from exam catalogues in primary school for the period 1936-1947 which is linked to birth record data. The matching was based on an algorithm using birth parish, date of birth, forenames and surnames.[1] Our purpose

[1] The data collection and digitisation is still in progress. Currently the match rate is roughly 50% but it is expected to increase up to about 80-90 % when all school data information is fully digitalised. As can be seen from the difference in means in Table 3 in the appendix, the sample is not complete. Thus, this is a very preliminary analysis.

built dataset consists of 10,900 individuals and includes birth information, several background characteristics (e.g. parental socio-economic status), information on sickness absence and school performance in math, reading and speaking, writing, Christianity and sports for grade 1 and 4.

The intervention included components delivered to mothers before giving birth and components delivered to infants. We will mainly focus on the effects of the infant intervention although we also control for being eligible for the maternal intervention. Having exact information on the date of birth of the individual and on rules for programme eligibility, we can determine the amount of months an individual was exposed to the intervention. We then estimate the impacts of exposure duration in treated in comparison to control regions applying difference-in-differences (DID).

Our estimation results suggest that the infant intervention had a significant effect on cognitive measures for boys. These results for boys are mainly due to an increase in academic performance in 'writing' and 'reading and speaking' in grade 4, whereas we find no improvements in grade 1. The coefficients for sickness absence are large in the magnitude of 0.15 days less absence per month of eligibility, but insignificant. The effects for the improvement of our cognitive grade-point average are rather small but one has to keep in mind that this is the effect per month of eligibility. Boys in grade 4 experienced an increase of 0.01 points on our grading scale per month of eligibility which equals about 2 % of a standard deviation. For comparison, the male/female gap in grade 4 equals 0.33 points.

The rest of the paper is structured as follows: Sections 2 and 3 give some background information on the field trial and the educational system during our investigation period; Section 4 explains our dataset and the matching procedure; results of the infant intervention are given in Section 5 and robustness checks are presented in Section 6.

2 The Field Trial

Sweden, just like many other developed countries, experienced a decline in maternal and infant mortality at the beginning of the 20th century, but in the 1920-1930 period no further improvement in infant mortality took place. This stagnation, together with increasing maternal mortality rates from 1920 onwards, and an increasing number of deaths of children occurring during the first few days of life, gave rise to intense public debate in Sweden how to improve conditions for expectant mothers and newborns.[2] The solution suggested was the field trial we analyse in our paper.

The field experiment started on 1 October 1931 and ended in June 1933. It involved 7 health districts (Lidköping, Hälsingborg, Harad, Råneå, Jokkmokk, Pajala and Mörtfors) which received free ante- and neonatal care during this period. The goal was to choose health districts that together mirrored the situation in the country as a whole. Importantly, the districts were only chosen based on variables like population density and living standards to reflect the diversity in local conditions and not due to the outcome variables (i.e. school marks and sickness absence) we analyse. The aim of the intervention was to modernise maternal and infant care and involved funding of SEK 30,000 (USD 133,000 in current prices) in order to finance the activities to meet this purpose.

To ensure uniform standards, a 5 day long educational event including lectures and courses for the participating staff was organised in Stockholm in July 1931. One of these standards demanded that doctors and nurses had to keep records of all activities they exerted, using utilisation forms for every child and mother who enrolled. In each of the 7 districts a health center was started and outreaching activities to inform the population about the existence of the services took place.

The main activities focused on preventive care and included guidance services and examinations at the surgeries, home visits and information campaigns. Beside of this, the maternal care provided anamnesis, antenatal examinations like temperature testing or the analysis of haemoglobin and urine samples and gave dietary recommendations. The antenatal care furthermore included recommendations for outdoor exercise, personal hygiene and sex and general information on maternal and infant care was provided by the National Board of Health. The visits to the health centres when the child was born allocated further check-ups and services like the weighting of the child and the referral of sick children to the doctor. Mothers were encouraged to breastfeed their children and with the neonatal care they received written and illustrated details on adequate nutrition of the child at its various stages of development. When the baby was born nurses made home visits aimed at giving advice on hygiene, sanitation and cleanliness in the household. Furthermore, these visits aimed at ensuring that the families followed the recommendations they were given and at getting an idea of the environment and circumstances

[2] The situation with worsening maternal mortality was not unique to Sweden. Rather maternal mortality did not decrease in any Western country between 1920 and the mid 1930's. The reasons for this development are still unclear, but some parts of the negative trend might be ascribed to higher rates of septic complications following higher rates of abortions.

the child was exposed to.

The eligibility of the programme was determined by birth date. Children less than one year of age were eligible at the beginning of the trial and children born during the intervention became eligible. Expectant mothers were authorised to participate during the whole time of the intervention, irrespective of their stage of pregnancy.[3] Figure 1 shows the duration of eligibility in months for the maternal and infant intervention by birthdate. The vertical lines represent the beginning and the end of the trial. The project was well received in the test districts. The health centres received on average 2 visits per mother and about 0.9 received a home call. The response to the child intervention was somewhat higher with 2.8 visits and 3.9 home calls on average per child.

[**Insert Figure 1 about here**]

The trial was so positively evaluated that a similar scheme was rolled-out nationwide in 1937. Physicians as well as auditing reports (Stenhoff, 1934) attributed improvements in maternal and infant health, respectively behavioural changes among participants to the intervention. For example the audit report from the chief physician of the northern districts of Sweden states that there was a notable change in the cleanness and tidiness of childrens beds and clothing. In total about 2,000 mothers and 2,600 children enroled which represents a majority of eligible individuals. In general, the take-up in the two urban areas was lower than in the rural areas and the utilisation of the infant intervention was slightly higher than that for the maternal intervention. For detailed information on the field experiment and the utilisation of the services see Bhalotra et al. (2014).

3 The Swedish Educational System

In the 1930's schooling in Sweden started in the year one turned 7 and was compulsory for 6 years. The attendance rate was almost 100%. Pupils went to *Folkskolan* during these years and most of them ended their educational career after this time. Secondary schooling became more popular and widely spread geographically in the 1950s. From 1918 onwards, pupils that did not continue to secondary school were obliged to take two more years of vocational courses, but these were taught with very low intensity.

[3] Although these rules were not strictly enforced regarding the maximum eligibility of 12 months for children and the areal restrictions, there are only very few exceptional cases which do not threat the reliability of our results.

In 1919 a central education plan, the so called *utbildningsplanen* was introduced to overcome the variations and differences in content and format of primary schools in the different school districts.[4] These guidelines were published by the "Department of Ecclesiastical Affairs" and included time-tables, syllabuses for compulsory schooling and defined the possible forms a school could have. The curriculum did not change between 1919 and 1950.

Although most students attended school full time, which would mean roughly eight months per year, some school districts in rural areas provided half time reading in order to meet the demands from the agrarian sector. But in the beginning of the 1940s half time reading was only somewhat more important in the regions of *Halland* (7.6% of pupils), *Gothenburg and Bohus* (15.3%) and *Älvsborg* (24.1%). In our sample half time reading accounts for less than 0.5%. Beside of the possibility of having half time or full time reading, *Folkskolan* was divided into two groups: *The Main forms* and *The Exception forms.* The main forms consisted of the forms A and B. Both of these required full time reading and an appropriate teacher degree (*folkskollärareexamen*) of the teacher but in comparison to the B-forms, where one teacher supervised several grades, each teacher in the A-forms supervised only one grade. The exception forms were defined by the C and D-forms. In comparison to the main forms, these were either characterised by half time reading (C and D) or by the fact that the teacher did not have an appropriate degree to teach in *folkskolan* (only D-forms). The exception forms were only accepted if the local conditions allowed for no other forms. That is why at the beginning of the 1940s, more than 90% of all pupils in Sweden went to a school which can be assigned to the main forms. An overview on the proportion of the different school forms in the school year 1940/1941 in comparison to the proportions in our sample is provided in Table 1.

[Insert Table 1 about here]

Teachers recorded marks and sickness absence (among other things) of all children in exam catalogues. The grading scale used throughout the period was introduced in 1897, and was applicable to all subjects but not to behavioural marks (these had a shorter scale and much higher concentration in the highest marks). Officially marks were given on a seven-point grading scale which ranged from A (passed with great distinction) to C (failed). Teachers were also allowed to use + and - signs to express the strength or weakness of a mark. A complete list of applied marks and their meaning can be seen in Table 2. At the outset, there was some heterogeneity in how student performance was evaluated, but our investigation period falls during a period

[4]The country was divided into about 2,400 school districts during this time.

of constantly increasing comparability between schools and teachers in their marking of pupil performance.

[Insert Table 2 about here]

A pass mark, i.e. at least a B, was required in theoretical subjects to proceed to the next grade.[5] There was, however, some local variation in how this rule was enforced in practice: some districts required a pass mark in all theoretical subjects; some allowed for a maximum of two fails, provided these two are not Swedish and math. Other districts allowed for very high marks in some subjects to offset low marks in other subjects.

Since from 1939 onwards, admission to secondary school was based on marks from primary school, a Royal Commission emphasised that the marking procedure should be improved and standardised much more. Therefore, guidelines for marking were prepared which were published in 1940 and became official starting with the school year 1940/41. These provided general guidelines for the marks and gave further information on individual subjects. It was stated that marks should be defined in a relative sense, meaning that *Ba* is defined as the normal mark which should encompass the middle third of a pupil's cohort. Consequently, one third of the other pupils should fall below this mark and the other third should be above. Only in really exceptional cases pupils obtained the extreme marks C or A. According to the commission, less than one percent of the pupils could be expected to have the knowledge corresponding to the top mark A, which should testify exceptional talent.

In order to ensure high reliability and comparability of the marks, several other principles were established. For example teachers were asked to mark the quality of knowledge and not the quantity and to take notes throughout the year. This ensured that grading didn't only take place retrospectively at one point in time. Thus, the marks we analyse differ from regular test scores which also depend on the pupil's form on the day of the test. Marks should therefore reflect the general knowledge of the students quite well. Furthermore, teachers should not allow for mark inflation as pupils progress to higher grades and they were told that school form should not be adjusted for. According to the guidelines, the main forms (A and B) should be marked according to the same principles. For C schools a different marking was allowed, but since these were so rare it is hardly an issue. Nevertheless we control for school form in our analyses. From 1943/44 onwards, the national school authorities explicitly recommended

[5]There are only very few statistics on how common grade retention was at that time, but a survey in 1940 from the second largest city of Sweden Gothenburg suggests that about 3% of all pupils had to repeat a grade (Paulsson, 1946)

teachers to use their standardised tests in Swedish and math to anchor the final marks each year. All these measures indicate that standardisation and reliability of school marks was a big issue during our investigation period. The reliability of the marks will be confirmed in a robustness check in Section 6. So far it can be stated that marks seem to be informative, although the most recent ones are probably more reliable than the older ones which were given before the recommendations took place. Since we observe marks between 1936-1947, this statement holds true for our sample as well. We will validate our results anchoring the grading scale to adult income in 2002.

Beside of these marking instructions, the Swedish educational system experienced a lot of changes and reforms concerning compulsory schooling and the length of the school year from 1936 onwards. These changes were largely quantitative whereas the system itself remained unchanged. We are able to control for *all* of these changes in our analyses. For further details on these reforms see Fischer et al. (2013).

4 Data and Empirical Strategy

4.1 Data Sources

The dataset we are using is unique and gives us various exceptional opportunities. Our main dataset was developed by merging two different individual level datasets; both purpose-built for analysing the effects of early life health interventions. The first dataset contains *individual-level data from all births* in the identified treated and control regions for the cohorts of 1930-1934. Since the trial took place in seven medical districts which consisted of 2 cities and 57 rural parishes, we identified 2 control cities and 57 control parishes (belonging to 38 different health districts) using observable parish characteristics of the 1930 census. Figure 2 visualises these areas on the municipality level. The best matches in terms of observable characteristics (denoted $\mathcal{J}_M(i)$) were identified using the Mahalanobis distance metric, defined as

$$\mathcal{J}_M(i) = \arg\min_j \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)} \tag{1}$$

where X_i is a vector of observable characteristics for a parish belonging to a test district, in our case, average income; net wealth; employment shares in manufacturing and agriculture; population density; proportion of fertile married women; and a dummy variable for urban locations, and S denotes the covariance matrix of the vector of observable characteristics. Since the matching was done before the data collection took place, it does not take into account our outcome variables. In fact, our matching procedure is based on similar information available to the decision makers of the time of the intervention. The matching was done in random order and without replacement. For further information on the identification of the control group and the underlying matching procedure see Bhalotra et al. (2014).

[Insert Figure 2 about here]

The information in the first dataset comes from high-quality administrative church records on all newborns. Sweden is one of few counties with high-quality vital statistics at the parish level covering the universe of the population from the 18th century. In our treatment and control regions there were in total 24,710 deliveries (25,029 individual children) during this time which resulted in 24,373 births. For all these individuals we observe various background characteristics like children's sex, marital status, parental profession and mother's age. We also observe paternal occupational status which was translated into occupational classes based on the HISCO classification (Leuwen et al., 2002), allowing us to control for socio-economic status. After the collection of the main data, we added information from several other data sources via matching procedures which were carefully executed and validated (cf. Bhalotra et al., 2014). We therefore also have information on death dates of the individuals which result from merging with the Swedish Death Index (cf. Fischer et al., 2013) and income information for the years 2002-2013 from tax records.

The second dataset contains information on *school performance and sickness absence*[6] for a subsample (the digitisation is still in progress) of the treated and control regions for the school years 1936-1947. It was gathered by digitising standardised exam catalogues from historical archives which contain individual information on each pupil in each school year. We mainly observe our cohorts in *folkskolan* in grade 1 and grade 4. The reason for not collecting data on later grades is that children continuing to secondary schooling could chose to do so after grade 4 (or after grade 6). For about half of our sample we have details from both grade 1 and grade 4. The other half of the sample is only observed once; either in grade 1 or in grade 4 which likely can be explained by the ongoing data collection. Due to the possibility of grade retention we rarely observe pupils more than once per grade. Detailed information on school performance is available for five subjects: math, writing, reading and speaking, christianity and sports. Besides

[6]Sickness absence accounts for about 80% of total absence.

this, the school dataset also contains information on sickness absence, total absence in days, the length of the school year, school type or the name of the teacher and the name of the school.

In order to analyse the effects of the maternal and child health intervention on academic performance, the individuals in the birth records were matched to the school data via an algorithm based on birth parish, date of birth, forenames and surnames. Out of 22,500 individuals still alive until age 7 from the 1930-1934 cohort, roughly 10,900 were matched to our school data. The left column of Table 3 presents summary statistics for the main explanatory variables in our matched sample.

[Insert Table 3 about here]

Although the data digitisation is still in progress, we can assume that the unmatched individuals will be missing at random once all data has been typed. Selection, due to migration to another parish after birth, should be unproblematic since the matching algorithm already takes this into account for movement between treatment and control regions. Currently we are gathering data on academic performance on individuals who moved to places which were not defined as treated or matched control and therefore were not covered in the digitisation progress and the matching algorithm. Future results should therefore be quite reliable although at the moment we can only assume that the unmatched individuals are unmatched at random. Thus, this is a preliminary analysis of the effects of the field trial on schooling and health.

Table 4 shows descriptive statistics and p-values for the differences in means for the outcome variables in our treatment and control regions before, during and after the intervention. Although treatment and control group seem to differ in sickness absence and sports before the intervention took place, the differences in means are insignificant concerning our cognitive measurements. Furthermore, these differences should not be problematic since we also control for health district specific differences and health district specific time trends later on. As will be shown later in Section 6, we also conduct tests for pre-trends.

[Insert Table 4 about here]

4.2 Empirical strategy

Our main focus is on the effects of the infant intervention on sickness absence (which is measured in days) and academic performance. We use difference-in-differences estimation to compare our exposed cohorts in the treated regions to our control group. Since marks were officially given on a 7-point grading scale in the 1930s, we translated the scale into a range from 1 for the poorest mark (C) to 7 for the best mark (A). An adjustment of plus respectively minus .3 was made if the original recorded marks contained a plus or a minus. As an example the mark *a+* would get a value of 6.3 whereas *a-* would be translated into 5.7. That way we retain the variation in the variable which is given by the indication of strength and weakness of a special mark. We create a measure of children's cognitive abilities by taking the mean of school performance in 'math', 'reading and speaking' and 'writing' but we also look at the subjects one by one. The effects are separately estimated for grade 1 and grade 4 and we also conduct the analysis separately for the subgroups of males and females. Figures 3 and 4 show the fraction of the grade point average by gender and grade. Apparently females got better marks than males and marks in grade 4 had a higher variation and improved in comparison to grade 1.

[**Insert Figure 3 about here**]

[**Insert Figure 4 about here**]

Since marks are given on an ordinal scale and thus only allow to define ranks instead of a cardinal interpretation, the choice of the scale can have significant implications for our findings (cf. Bond and Lang, 2013; Lang, 2010). To address this issue we will anchor the 7-point grading scale to income in adulthood (as proposed e.g. by Cunha and Heckman (2008) in a robustness check (see Section 6). Applying this method gives us an interpretable metric and a valid scale based on labour income in 2002. We first gauge the overall impact of the baby intervention:

$$y_{ipt} = \alpha + \beta_I T_{It} + \gamma D_p + \delta_I T_{It} D_p + \varepsilon X + u_{ipt}$$

where y_{ipt} is the outcome of child i born in health district p on day t, T_{It} is the duration of eligibility for the infant intervention I for a child i born on day t in months, D_p is a dummy for treated districts and X is a vector of covariates. Covariates that we condition on include the sex of the child, marital status of the mother, a twin indicator and dummies capturing old (>35 years) and young (<25) mothers. Furthermore, we also control for the effects of the maternal intervention. The outstanding variety of information in our datasets also allows us to control for the length of the school year, the school form (an indicator of school quality) and parental occupational status. We also include school fixed effects, $Quarter\ of\ birth \times Year\ of\ birth$ ($QOB \times YOB$) effects and health district fixed effects to validate the results. In order to allow for a possible different trend between treatment and control regions, we add health district specific time-trends in our last specification which is more general than treatment-group-specific trends.

Assuming a common time trend in the absence of the treatment for treatment and control regions, the parameter δ_I measures the intent-to-treat (ITT) effect of the infant intervention. It gives us the effect for each additional month of eligibility of making the service available. This effect is especially of importance for policy makers who are unable or unwilling to make the utilisation of services mandatory.

5 Results

5.1 Academic Performance

Our first set of estimation results explores the effect of the infant intervention on children's cognitive abilities calculating the grade-point average (GPA) as described in Section 4. Table 5 displays the difference-in-differences estimation results for all possible gender-grade combinations. All specifications control for our before mentioned set of covariates (see Table 3). We identify significant effects on marks for males which is clearly driven by their improvement in marks in grade 4. Although small in magnitude, one has to keep in mind, that this is the effect for one more month of eligibility. Thus, being eligible for one more month improved the cognitive GPA for males by about 0.007 points on the grading scale. If a child would have been eligible during the complete intervention period, this would mean an improvement of roughly 0.084 points. These results are robust to the inclusion of health district specific time trends.

[Insert Table 5 about here]

We do not find any effects of the infant intervention on our cognitive GPA measure for girls and boys in grade 1. The missing effect in grade 1 could be due to the guidelines of the Royal Commission (see Section 3). These stated that in the first grade, the grading scale shall be used with some caution and extreme marks shall be avoided. Thus, the marks from grade 4 are probably more reliable than those of grade 1. Furthermore, grade 4 marks were used in the admission to secondary school and thus had a more official status than those of grade 1. Anchoring of the grading scale to later life income in 2002 will solve these issues.

We also explore the subjects separately; once again for each gender-grade combination. Since health district specific time trends demand a lot from our data and because results seem to be quite robust to their inclusion, all of the estimates for the subjects refer to specification (7) of

Table 5 which includes all controls except for health district specific time trends. As can be seen in Table 6, the results for males in grade 4 of our cognitive GPA are mainly driven by improvements in 'writing' and 'reading and speaking', and there is no significant improvement in math. No effects are visible for Christianity which is not surprising assuming that the marks in these subjects are more likely to be achieved due to inheritance of religious skills from the parental home rather than effort. Surprisingly the infant intervention seems to have reduced the performance of girls in sports in grade 1 and grade 4. Girls which were exposed to the intervention one more month have on average 0.008 lower points on the grading scale than those who were not exposed. We will have to explore this further but a possible explanation could be endogenous survival selection as, post-intervention, the marginal survivor was negatively selected. Bhalotra et al. (2014) show that the intervention lowered infant mortality by around 1 percentage point on average from a base rate of 6.6 percent. If selection plays a significant role for girls in sports, it potentially plays a significant role for boys and in other domains. In the presence of selection bias, the estimated impact of the intervention on boys' GPA is an under-estimate.

[Insert Table 6 about here]

5.2 Sickness absence

The infant intervention seems to have had no significant effect on sickness absence (measured in days). Although relatively large in magnitude (0.15 days less sickness absence in grade 1), the significance from the first specification without any controls is lost, once we add further controls. The results of the intervention are displayed in Table 7. Nevertheless, the results in grade 1 seem to be driven by girls and due to the ongoing data collection it is possible that the results turn out to be significant, once the data collection is completed.

[Insert Table 7 about here]

6 Robustness checks

6.1 Common trend assumption

To investigate whether the treatment and control regions follow a similar trend, we test for pre-trends, looking at our significant outcome variables before the intervention took place. We run regressions for the pre-intervention period based on the following equation:

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

$Trend$ is a trend variable based on each $month \times year$ observation in our pre-intervention sample and *treated* refers to a dummy indicating treated parishes. If β is unequal to 0, the treated and control regions had a significantly different pre-trend. Table 8 indicates, that this is not the case.

[Insert Table 8 about here]

Furthermore our results are quite robust to the inclusion of health district specific time trends. These findings make it quite likely that treatment and control region didn't experience a significantly different pre-trend.

6.2 Anchoring of grading scale

Our findings regarding the positive effects for males in 'writing' and 'reading and speaking' in fourth grade could in fact be sensitive to the choice of the applied grading scale (Bond and Lang, 2013). In order to validate our results on academic performance and to confirm the reliability of the 7-point grading scale, we anchor the scale to labour income in 2002. Although the school marks are given on an ordinal scale, we will show that they are interpretable by anchoring them to the cardinal scale of labour market income in adulthood, as proposed e.g. by Cunha and Heckman (2008). In comparison to their method we use labour income in levels instead of logarithmised labour income.

We separately regress labour income in 2002 on the 7-point grading scale for each subject in each grade. Thus, we obtain the mean income in SEK for each step on our grading scale. Tables 9 and 10 list the results for the the cognitive subjects of 'math', 'reading and speaking' and 'writing' in grade 1 respectively grade 4. Since labour income increases with better marks in grade 1 and grade 4, the grading scale can perfectly be interpreted and seems to be quite informative and reliable. In a next step we will use the underlying anchored scales to validate our results for academic performance in each subject. For the case of our cognitive grade point average we will apply a non-parametric approach to overcome the problem of possible school mark correlations.

[Insert Table 9 about here]

[Insert Table 10 about here]

7 Conclusion

Our very preliminary analysis indicates that an improvement in early life health conditions through basic neonatal care leads to better academic performance in primary school. This is in line with e.g. Cunha and Heckman (2007) or Almond and Currie (2011) who state that the early years are a crucial period for later life outcomes and cognitive development. Exploiting a Swedish field trial in the 1930s, which improved infant care, as a source of exogenous variation in health, we show that males experienced an increase in cognitive GPA which is mainly driven by marks in grade 4. The effect per month of eligibility is 0.01 points on our grading scale which equals about 2 % of a standard deviation. We find no significant effects on academic performance in grade 1, which represents an earlier key stage of children's cognitive development. The effects in fourth grade are mainly due to improved marks in 'writing' and 'reading and speaking'. Although large in magnitude (0.15 days less absence in grade 1 per month of eligibility), we find (at the current state of data collection) no significant effects of the infant intervention on sickness absence. Nevertheless, this effect seems to be clearly driven by females.

Taking into account these findings on school marks and the before produced results of Bhalotra et al. (2014) on mortality, the Swedish infant intervention of 1931-1933 seems to have had a large return on the investment. Estimates by Grantham-McGregor et al. (2007) suggest that poverty, poor health and nutrition, and deficient care will cause about 200 million children under the age of 5 not to reach their full potential in cognitive development. This disadvantage can in turn give rise to poor levels of cognition and education later in life and even to the intergenerational transmission of poverty due to reduced wages and income (Grantham-McGregor et al., 2007). Improving basic infant care could therefore result in better health and academic achievement in todays low income countries, which in turn could lead to reduced inequalities.

References

Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in-utero influenza exposure in the post-1940 u.s. population. *Journal of Political Economy 114*(4), 672–712.

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

Almond, D., L. Edlund, and M. Palme (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly Journal of Economics 124*(4), 1729–1772.

Behrman, J. R. and M. R. Rosenzweig (2004). Returns to birthweight. *The Review of Economics and Statistics 86*(2), 586–601.

Bhalotra, S., M. Karlsson, and T. Nilsson (2014). Life expectancy and mother-baby interventions: Evidence from a historical trial. *Working Paper*.

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

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.

Black, S. E., P. J. Devereux, K. G. Salvanes, and A. Bütikofer (2013). This is only a test? long-run impacts of prenatal exposure to radioactive fallout. *NBER Working Paper Series*.

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.

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

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

Currie, J. and M. Rossin-Sater (2014). Early-life origins of lifecycle wellbeing: Research and policy implications.

Currie, J. and T. Vogl (2013). Early-life health and adult circumstance in developing countries. *Annual Review of Economics 5*(1), 1–36.

Del Bono, E. and B. Rabe (2012). Breastfeeding and child cognitive outcomes: Evidence from a hospital-based breastfeeding support policy. *ISER Working Paper Series* (2012-29).

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), 3921–103955.

Fischer, M., M. Karlsson, and T. Nilsson (2013). Effects of compulsory schooling on mortality: evidence from sweden. *International journal of environmental research and public health 10*(8), 3596–3618.

Fuchs, V. R. (1982). Time preference and health: An exploratory study. *Economic Aspects of Health*, 93–120.

Glewwe, P., H. G. Jacoby, and E. M. King (2001). Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of Public Economics 81*(3), 345–368.

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.

Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy 80*(2), 223–255.

Lang, K. (2010). Measurement matters: Perspectives on education policy from an economist and school board member. *Journal of Economic Perspectives 24*(3), 167–82.

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

Paulsson, E. (1946). Om folkskoleväsendets tillstånd och utveckling i sverige under 1920- och 1930-talen. *Jönköping: Länstryckeriaktiebolaget*.

Razavi, S. (2012). World development report 2012: Gender equality and development - a commentary. *Development and Change 43*(1), 423–437.

Royer, H. (2009). Separated at girth: Us twin estimates of the effect of birth weight. *American Economic Journal: Applied Economics 2009 1*(1), 49–85.

Sanders, N. J. (2012). What doesn't kill you makes you weaker: Prenatal pollution exposure and educational outcomes. *Journal of Human Resources 47*(3), 826–850.

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

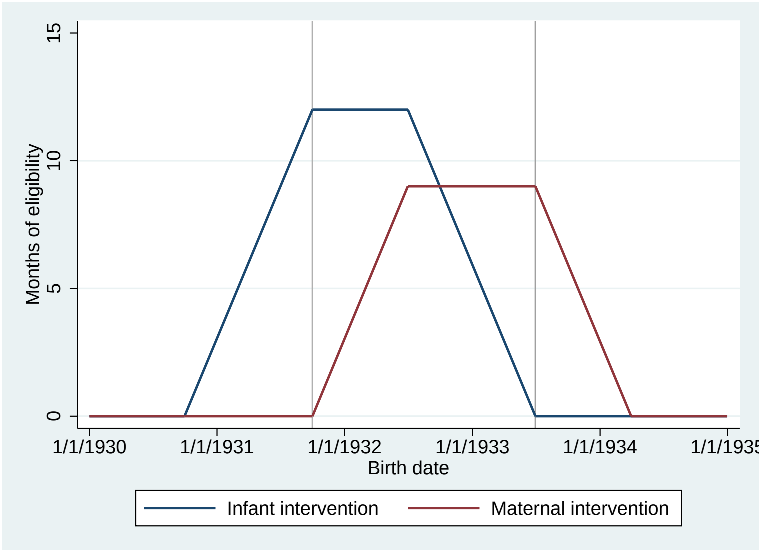

Figure 1. Eligibility by birth date

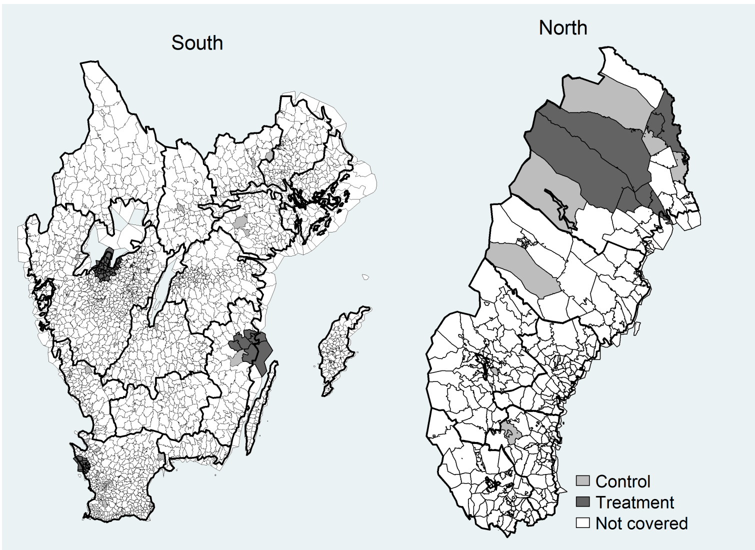

Figure 2. Municipalities containing treated and control districts

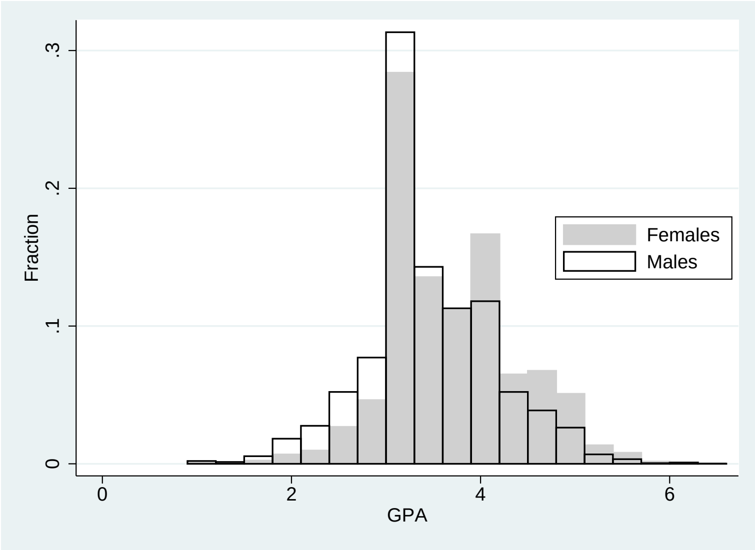

Figure 3. Fraction of GPA by gender

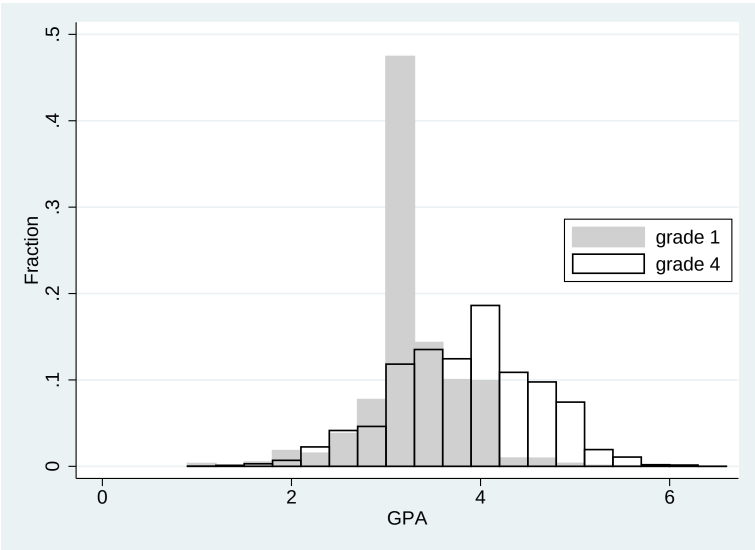

Figure 4. Fraction of GPA by grade

Table 1. School form

	Form	Sample	1940/1941
Full time reading	A	36.97%	44.9%
	B1	38.01%	26.4%
	B2	17.14%	19.2%
	B3	2.59%	3.3%
	D1	4.81%	2.5%
	aid-class	-	1.4%
Half time reading	C1	0.40%	1.8%
	C2	-	0.3%
	D2	0.09%	0.2%
	D3	-	0.0%

Table 2. 1897 grading scale

Mark	Name	English
A	*Berömlig*	Passed with great distinction
a	*Med utmärkt beröm godkänd*	Passed with distinction
AB	*Med beröm godkänd*	Passed with great credit
Ba	*Icke utan beröm godkänd*	Passed with credit
B	*Godkänd*	Passed
BC	*Icke fullt godkänd*	Not entirely passable
C	*Underkänd*	Fail

Table 3. Descriptive statistics

	Matched individuals					**Unmatched individuals**					Difference in means
	count	mean	sd	min	max	count	mean	sd	min	max	p-value
Female	10,938	0.495	0.500	0	1	11,616	0.488	0.500	0	1	.3366405
Wedlock	10,938	0.926	0.262	0	1	11,616	0.873	0.333	0	1	1.42e-39
Twin	10,938	0.024	0.153	0	1	11,616	0.022	0.146	0	1	.2374439
Momfinnish	10,938	0.074	0.261	0	1	11,616	0.095	0.294	0	1	4.20e-09
Dadfinnish	10,938	0.064	0.244	0	1	11,616	0.079	0.270	0	1	6.80e-06
Treated	10,938	0.550	0.498	0	1	11,616	0.578	0.494	0	1	.0000226
Mother<20	10,938	0.042	0.200	0	1	11,616	0.060	0.238	0	1	3.03e-10
Mother>35	10,938	0.236	0.425	0	1	11,616	0.210	0.407	0	1	1.97e-06
SES_Professional/Technical	10,938	0.038	0.192	0	1	11,616	0.055	0.228	0	1	3.26e-09
SES_Administrative/Managerial	10,938	0.017	0.128	0	1	11,616	0.025	0.156	0	1	.0000192
SES_Clerical	10,938	0.011	0.106	0	1	11,616	0.019	0.137	0	1	2.22e-06
SES_Sales	10,938	0.024	0.154	0	1	11,616	0.028	0.165	0	1	.0654917
SES_Service	10,938	0.012	0.108	0	1	11,616	0.025	0.155	0	1	1.02e-12
SES_Agricultural	10,938	0.396	0.489	0	1	11,616	0.318	0.466	0	1	2.26e-34
SES_Production	10,938	0.417	0.493	0	1	11,616	0.402	0.490	0	1	.0249269
SES_Unknown	10,938	0.085	0.279	0	1	11,616	0.128	0.334	0	1	6.32e-26
DurationI	10,938	4.266	4.814	0	12	11,616	4.193	4.822	0	12	.2544375
DurationM	10,938	3.066	3.781	0	9	11,616	3.114	3.787	0	9	.3409813
Duration	10,938	7.332	7.005	0	21	11,616	7.307	7.056	0	21	.7883108

Table 4. Descriptive statistics and p-values for difference in means

Before

	control					treated					p-value
	count	mean	sd	min	max	count	mean	sd	min	max	
sickn. absence	1,191	8.864	13.255	0	186	1,511	10.940	15.412	0	206	.0002248
writing	958	3.505	0.818	1	6	1,229	3.454	0.817	1	6	.1458164
readandspeak	1,195	3.624	0.757	1	7	1,493	3.613	0.725	1	6	.6945365
sports	1,098	3.410	0.579	2	7	1,434	3.355	0.559	2	5	.0165834
math	1,193	3.509	0.808	1	6	1,493	3.553	0.842	1	6	.1701572
christianity	1,185	3.489	0.790	1	6	1,488	3.467	0.750	1	6	.460432
GPA_cog	1,195	3.528	0.703	1	6	1,495	3.527	0.698	1	6	.9741133

During

	control					treated					p-value
	count	mean	sd	min	max	count	mean	sd	min	max	
sickn. absence	5,164	9.363	13.285	0	217	6,290	10.375	13.387	0	205	.0000537
writing	3,921	3.567	0.852	1	6	5,463	3.488	0.830	1	6	7.33e-06
readandspeak	5,166	3.635	0.766	1	6	6,248	3.636	0.766	1	7	.9466921
sports	4,655	3.432	0.630	1	7	6,108	3.396	0.600	1	6	.0022145
math	5,168	3.530	0.847	1	7	6,244	3.542	0.862	1	7	.450067
christianity	5,141	3.494	0.779	1	7	6,220	3.494	0.776	1	6	.9695523
GPA_cog	5,172	3.550	0.720	1	6	6,248	3.544	0.723	1	6	.6550464

After

	control					treated					p-value
	count	mean	sd	min	max	count	mean	sd	min	max	
sickn. absence	1,079	9.391	11.578	0	125	1,233	10.442	12.275	0	118	.0350975
writing	799	3.521	0.822	1	6	1,094	3.480	0.815	1	6	.2892183
readandspeak	1,082	3.603	0.759	1	6	1,231	3.629	0.766	1	6	.4125768
sports	1,011	3.370	0.590	1	6	1,213	3.411	0.586	2	6	.1060981
math	1,083	3.494	0.784	1	6	1,231	3.512	0.843	1	6	.6019556
christianity	1,082	3.446	0.733	1	6	1,228	3.512	0.776	1	7	.0353911
GPA_cog	1,083	3.515	0.699	1	6	1,231	3.531	0.712	1	6	.5889078

Table 5. DID estimates of the infant intervention for GPA_cog

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Males and females	**grade 1**	0.0011	0.0008	0.0010	-0.0006	-0.0005	-0.0007	-0.0010	-0.0028
	SE	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	N	8,326	8,326	8,324	8,324	8,324	8,321	8,321	8,321
	ymean	3.247	3.247	3.247	3.247	3.247	3.247	3.247	3.247
	grade 4	0.0063	0.0059	0.0061	0.0067	0.0075	0.0074	0.0076	0.0077
	SE	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.005
	N	8,098	8,098	8,097	8,097	8,097	8,090	8,090	8,090
	ymean	3.842	3.842	3.842	3.842	3.842	3.842	3.842	3.842
Grade 1 and grade 4	**males**	0.0065	0.0064	0.0068	0.0076*	0.0078*	0.0074**	0.0077**	0.0076*
	SE	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.004
	N	8,337	8,337	8,335	8,335	8,335	8,330	8,330	8,330
	ymean	3.421	3.421	3.421	3.421	3.421	3.422	3.422	3.422
	females	0.0011	-0.0006	-0.0007	-0.0019	-0.0013	-0.0003	-0.0002	0.0003
	SE	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.006
	N	8,087	8,087	8,086	8,086	8,086	8,081	8,081	8,081
	ymean	3.663	3.663	3.663	3.663	3.663	3.663	3.663	3.663
Males	**grade 1**	0.0038	0.0041	0.0046	0.0034	0.0032	0.0030	0.0025	0.0002
	SE	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.004
	N	4,294	4,294	4,293	4,293	4,293	4,292	4,292	4,292
	ymean	3.181	3.181	3.181	3.181	3.181	3.181	3.181	3.181
	grade 4	0.0071	0.0057	0.0062	0.0092	0.0101	0.0105	0.0107*	0.0101
	SE	0.005	0.006	0.005	0.006	0.006	0.006	0.006	0.007
	N	4,043	4,043	4,042	4,042	4,042	4,038	4,038	4,038
	ymean	3.676	3.676	3.677	3.677	3.677	3.677	3.677	3.677
Females	**grade 1**	-0.0020	-0.0031	-0.0027	-0.0034	-0.0032	-0.0034	-0.0034	-0.0053
	SE	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	N	4,032	4,032	4,031	4,031	4,031	4,029	4,029	4,029
	ymean	3.318	3.318	3.318	3.318	3.318	3.318	3.318	3.318
	grade 4	0.0054	0.0045	0.0040	0.0023	0.0037	0.0029	0.0033	0.0029
	SE	0.007	0.008	0.008	0.009	0.009	0.009	0.009	0.010
	N	4,055	4,055	4,055	4,055	4,055	4,052	4,052	4,052
	ymean	4.007	4.007	4.007	4.007	4.007	4.007	4.007	4.007
	Health District FE		✓	✓	✓	✓	✓	✓	✓
	QOB*YOB effects			✓	✓	✓	✓	✓	✓
	School FE				✓	✓	✓	✓	✓
	SES effects					✓	✓	✓	✓
	Length of Schoolyear						✓	✓	✓
	Schoolform							✓	✓
	Health District*Year effects								✓

Note: *** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the health district level. Covariates which are included in all specifications are a dummy indicating twin births, a dummy for being female, dummies capturing old (>35 years) and young (<20) mothers, a dummy for married women and the treatment effect of the maternal intervention. *ymean* refers to the mean value of the outcome variable. *QOB× YOB effects* include quarter-of-birth dummies for each of the 20 quarters. *Health District FE* are fixed effects for the health district 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 the reforms concerning the length of the school year. *Schoolform* are fixed effects controlling for the school form as described in Section 3 and *Health District× Year effects* allows for health district specific time trends.

Table 6. DID estimates for subjects

		math	reading	writing	christianity	sports
Males and females	**grade 1**	-0.0027	0.0014	-0.0007	-0.0027	-0.0039***
	SE	0.003	0.005	0.005	0.003	0.001
	N	8,311	8,313	5,365	8,242	7,571
	ymean	3.255	3.304	3.156	3.094	3.046
	grade 4	0.0052	0.0093*	0.0082	0.0040	-0.0050
	SE	0.007	0.005	0.005	0.004	0.004
	N	8,088	8,089	8,088	8,089	7,935
	ymean	3.816	3.966	3.746	3.892	3.745
Grade 1 and grade 4	**males**	0.0061	0.0083*	0.0083*	0.0026	-0.0004
	SE	0.004	0.004	0.005	0.004	0.004
	N	8,324	8,328	6,782	8,292	7,871
	ymean	3.468	3.500	3.320	3.414	3.407
	females	-0.0013	0.0013	-0.0001	-0.0016	-0.0079*
	SE	0.007	0.006	0.009	0.005	0.004
	N	8,075	8,074	6,671	8,039	7,635
	ymean	3.597	3.764	3.704	3.567	3.400
Males	**grade 1**	0.0015	0.0040	-0.0019	-0.0010	-0.0013
	SE	0.004	0.005	0.005	0.004	0.002
	N	4,288	4,290	2,744	4,254	3,895
	ymean	3.209	3.227	3.059	3.061	3.038
	grade 4	0.0059	0.0130**	0.0132*	0.0029	-0.0026
	SE	0.008	0.006	0.007	0.006	0.005
	N	4,036	4,038	4,038	4,038	3,976
	ymean	3.743	3.790	3.498	3.786	3.768
Females	**grade 1**	-0.0067	0.0002	0.0014	-0.0040	-0.0053**
	SE	0.005	0.005	0.006	0.004	0.003
	N	4,023	4,023	2,621	3,988	3,676
	ymean	3.304	3.386	3.257	3.129	3.054
	grade 4	0.0022	0.0047	0.0028	0.0026	-0.0091
	SE	0.011	0.007	0.010	0.007	0.006
	N	4,052	4,051	4,050	4,051	3,959
	ymean	3.888	4.140	3.993	3.998	3.722
	Health District FE	✓	✓	✓	✓	✓
	QOB*YOB effects	✓	✓	✓	✓	✓
	School FE	✓	✓	✓	✓	✓
	SES effects	✓	✓	✓	✓	✓
	Length of Schoolyear	✓	✓	✓	✓	✓
	Schoolform	✓	✓	✓	✓	✓
	Health District*Year effects					

Note: *** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the health district level. Covariates which are included in all specifications are a dummy indicating twin births, a dummy for being female, dummies capturing old (>35 years) and young (<20) mothers, a dummy for married women and the treatment effect of the maternal intervention. *ymean* refers to the mean value of the outcome variable. *QOB× YOB effects* include quarter-of-birth dummies for each of the 20 quarters. *Health District FE* are fixed effects for the health district 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 the reforms concerning the length of the school year. *Schoolform* are fixed effects controlling for the school form as described in Section 3 and *Health District× Year effects* allows for health district specific time trends.

Table 7. DID estimates of the infant intervention for sickness absence

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Males and females	**grade 1**	-0.1618*	-0.1393	-0.1424	-0.1824	-0.1827	-0.1557	-0.1638	-0.1553
	SE	0.090	0.095	0.092	0.117	0.117	0.121	0.121	0.145
	N	8,366	8,366	8,364	8,364	8,364	8,361	8,361	8,361
	ymean	10.755	10.755	10.756	10.756	10.756	10.760	10.760	10.760
	grade 4	0.0172	0.0288	0.0294	0.0358	0.0351	0.0337	0.0341	0.0519
	SE	0.065	0.064	0.063	0.077	0.077	0.077	0.076	0.068
	N	8,102	8,102	8,102	8,102	8,102	8,096	8,096	8,096
	ymean	9.100	9.100	9.100	9.100	9.100	9.102	9.102	9.102
Grade 1 and grade 4	**males**	-0.0788	-0.0711	-0.0732	-0.0653	-0.0662	-0.0682	-0.0670	-0.0517
	SE	0.073	0.077	0.076	0.084	0.083	0.081	0.083	0.089
	N	8,374	8,374	8,373	8,373	8,373	8,369	8,369	8,369
	ymean	9.498	9.498	9.499	9.499	9.499	9.500	9.500	9.500
	females	-0.0662	-0.0457	-0.0362	-0.0591	-0.0565	-0.0399	-0.0403	-0.0348
	SE	0.081	0.075	0.072	0.081	0.080	0.077	0.077	0.083
	N	8,094	8,094	8,093	8,093	8,093	8,088	8,088	8,088
	ymean	10.398	10.398	10.399	10.399	10.399	10.403	10.403	10.403
Males	**grade 1**	-0.1199	-0.1128	-0.1141	-0.0908	-0.0917	-0.0968	-0.1022	-0.0793
	SE	0.094	0.105	0.099	0.121	0.121	0.126	0.130	0.147
	N	4,325	4,325	4,324	4,324	4,324	4,323	4,323	4,323
	ymean	10.432	10.432	10.434	10.434	10.434	10.436	10.436	10.436
	grade 4	-0.0322	-0.0159	-0.0169	-0.0160	-0.0235	-0.0325	-0.0381	-0.0023
	SE	0.077	0.084	0.082	0.094	0.092	0.094	0.095	0.088
	N	4,049	4,049	4,049	4,049	4,049	4,046	4,046	4,046
	ymean	8.500	8.500	8.500	8.500	8.500	8.501	8.501	8.501
Females	**grade 1**	-0.2053	-0.1706	-0.1536	-0.2538*	-0.2503*	-0.1934	-0.1969	-0.1585
	SE	0.124	0.121	0.115	0.129	0.128	0.130	0.129	0.162
	N	4,041	4,041	4,040	4,040	4,040	4,038	4,038	4,038
	ymean	11.101	11.101	11.101	11.101	11.101	11.106	11.106	11.106
	grade 4	0.0673	0.0705	0.0732	0.1135	0.1186	0.1360	0.1437	0.1522
	SE	0.088	0.084	0.086	0.097	0.097	0.096	0.094	0.097
	N	4,053	4,053	4,053	4,053	4,053	4,050	4,050	4,050
	ymean	9.698	9.698	9.698	9.698	9.698	9.701	9.701	9.701
	Health District FE		✓	✓	✓	✓	✓	✓	✓
	QOB*YOB effects			✓	✓	✓	✓	✓	✓
	School FE				✓	✓	✓	✓	✓
	SES effects					✓	✓	✓	✓
	Length of Schoolyear						✓	✓	✓
	Schoolform							✓	✓
	Health District*Year effects								✓

Note: *** p <0,01; ** p <0,05; * p <0,1, Standard errors are clustered at the health district level. Covariates which are included in all specifications are a dummy indicating twin births, a dummy for being female, dummies capturing old (>35 years) and young (<20) mothers, a dummy for married women and the treatment effect of the maternal intervention. *ymean* refers to the mean value of the outcome variable. *QOB× YOB effects* include quarter-of-birth dummies for each of the 20 quarters. *Health District FE* are fixed effects for the health district 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 the reforms concerning the length of the school year. *Schoolform* are fixed effects controlling for the school form as described in Section 3 and *Health District× Year effects* allows for health district specific time trends.

Table 8. DID pre-trend test

	(1) GPA_cog	(2) GPA_cog males grade 4	(3) sickness absence	(4) sickness absence females grade 1	(5) reading	(6) reading males grade 4
trend × *treated*	0.0031	-0.0060	0.1056	-0.2075	-0.0008	0.0013
	0.011	0.021	0.219	0.461	0.011	0.021
trend	-0.0204***	-0.0318**	0.0041	0.3284	-0.0114	-0.0262*
	0.008	0.015	0.163	0.345	0.008	0.015
treated	-0.0204	-0.0168	1.5409	4.4946*	-0.0094	-0.0495
	0.061	0.123	1.263	2.641	0.064	0.123
_cons	3.6352***	3.8235***	8.8425***	6.8649***	3.6840***	3.8921***
	0.046	0.089	0.956	1.987	0.049	0.089
N	2,690	725	2,702	641	2,688	725

Note: *** p <0,01; ** p <0,05; * p <0,1, *trend* variable is based on *month* × *year* observations; *treated* refers to a dummy indicating treated parishes.

Table 9. Anchoring grade 1

	math	reading	writing
1 point	121,445***	128,502***	116,111***
	(17,253)	(17,568)	(21,104)
2 points	139,243***	138,506***	144,224***
	(6,035)	(6,509)	(6,544)
3 points	152,350***	152,295***	154,749***
	(1,441)	(1,488)	(1,496)
4 points	167,696***	164,732***	154,731***
	(2,451)	(2,253)	(3,218)
5 points	185,705***	185,429***	171,862***
	(9,634)	(9,072)	(10,922)
6 points	203,309***	195,175***	224,948***
	(29,381)	(20,286)	(28,898)
7 points	-	-	-
	-	-	-
r2	0.739	0.739	0.782
N	5,965	5,937	3,866

Note: SE in parenthesis, *** p <0,01; ** p <0,05; * p <0,1. The point values refer to the 7-point grading scale defined in Section 4. Higher values indicate a better grade on the grading scale.

Table 10. Anchoring grade 4

	math	reading	writing
1 point	124,957***	145,136***	141,002***
	(13,851)	(35,879)	(14,646)
2 points	140,554***	147,203***	146,380***
	(4,473)	(8,745)	(4,801)
3 points	142,861***	153,120***	154,467***
	(2,105)	(2,335)	(2,038)
4 points	153,961***	152,837***	153,778***
	(1,797)	(1,621)	(1,786)
5 points	180,247***	171,629***	170,643***
	(2,455)	(2,443)	(2,667)
6 points	216,305***	166,930***	188,604***
	(6,615)	(10,216)	(9,263)
7 points	-	153,856**	-
	-	(62,143)	-
r2	0.770	0.762	0.763
N	5,830	5,831	5,829

Note: SE in parenthesis, *** p <0,01; ** p <0,05; * p <0,1. The point values refer to the 7-point grading scale defined in Section 4. Higher values indicate a better grade on the grading scale.