Childhood Illness and School Performance in Denmark

Mille Bjørk and Katrine Søe[1]
(Insurance & Pension Denmark)

Katrine Søe[1]
(Insurance & Pension Denmark)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 4B-1: Disability Measurement and its Implications

WEDNESDAY, AUGUST 22, AFTERNOON

Childhood illness and school performance in Denmark*

Mille Bjørk and Katrine Søe[1]

July 2018

Preliminary draft, not to be quoted

Abstract

Research has shown a strong link between the incidence of severe illness and socioeconomic outcomes such as cognitive abilities, socioeconomic status and parental relationship. This paper examines the prevalence of severe childhood illnesses and its effect on school performance in a Danish context, with universal healthcare and access to detailed longitudinal patient register on national level. Danish register data enables us to observe every contact an individual has with the Danish hospital sector. Covering all children born between 1994-1999, we find 34,000 incidences of Danish children who experienced and survived a severe health condition until the age 14. Access to data of such high quality and coverage, enables us to conclude on effects from severe illness as a broad definition, and compare effects between diagnosis groups, e.g. effects on GPA from the incidence of cancer compared to effects from diseases of inner organs. We find that severe illness in childhood is an important and significant determinant for children's school performance and hence the child's ability to invest in human capital. We also find evidence for long run effects of severe childhood illness.

---

* We are grateful to many who have discussed this research with us and who have provided helpful comments and feedback, particularly Torben Tranæs, Bjarne Laursen, Palle Sørensen, Benn Duus, and Frederikke Munck. We thank Frederikke Hedstrøm for great research assistance.

[1] Insurance & Pension Denmark. Mille Bjørk, mbj@forsikringogpension.dk and Katrine Søe, kas@forsikringogpension.dk.

Introduction

A growing body of research finds that adverse health shocks in childhood may have long-term socioeconomic costs for children. Depending on the measurement, around 10 per cent of all children have suffered from a severe illness (Loft, 2011; Danish Health Authority, 2012; Currier et al., 2008). The consequences are extensive to both the child, its family and society at large. Some illnesses even seems to have permanent adverse effects on children's cognitive skills. Due to the hypothesis of Cunha and Heckman (2007) that "skills beget skill", meaning that children who fall behind in early school years may never catch up, the effect may persist all the way through the educational system and on to the labor market as well. It is therefore important to have a deep understanding of the long-term consequences that severe illness may bring on a child.

In a Danish context, this paper investigates how severe illness affects a child in terms of cognitive abilities, measured by school performance. We consider a broad selection of critical and chronic illnesses, and examine how they affect the grade point average (GPA) by the end of compulsory school in ninth grade. We use the comprehensive Danish individual level data that include detailed information on background, school performance and hospitalization of the Danish population. Because Denmark provides universal public healthcare, this study examines the link between childhood illness and GPA at the end of compulsory school in a setting that abstracts from differences in access to insurance coverage.

This study focuses on severe illness in a broad sense and includes severe somatic diseases with some minimum requirement of prevalence among children aged 0-14 who survives until the age of 17. We have selected a wide range of diagnoses, and this enables us to compare different types of diagnoses with respect to their effect on school performance. We aggregate the diagnoses into eight diagnosis groups (cancer, muscular dystrophy and arthritis, neurological disorders, diseases in the inner organs, hormonal and metabolic disorders, respiratory diseases, traumatic injuries and concussions).[2]

We follow all Danish children born in 1994-1999 from birth to the end of compulsory school. Through detailed individual records on hospitalization, we are able to identify children who were diagnosed with a disease included in the eight aggregated groups of diagnoses – including length of hospitalization and age at diagnosis. To estimate the medium to long run effects of childhood illness, we compare school performance of children, diagnosed from age 0-14[3], with children from the same birth cohorts who were never been diagnosed. We control for relevant background characteristics of the children and their parents, measured around the time of birth to avoid dependency in the explanatory variables.

Our empirical approach compares the school performance of children with severe illness to children without. The main results only include children who obtained a GPA and the average for all children in our population is estimated to be 6.5[4]. The results of this study indicate that children with a severe illness on average experience a 0.3 grade point reduction in GPA (4 per cent) compared to children with-

[2] Loft (2011) and Currie et al. (2008) considers similar diagnosis groups.
[3] We only include children who survives until age 17.
[4] The Danish grading system is a seven-point scale taking on values –3 (F), 0 (Fx), 2 (E), 4 (D), 7 (C), 10 (B) and 12 (A), where -3 and 0 are failed.

out illness. The largest effects come from neurological diseases, respiratory diseases, and traumatic injuries causing on average a 0.6 grade point reduction in GPA (9 per cent). We find that the negative effect of illness on GPA is increasing with severity, as measured by the length of hospitalization. We further find evidence for long run effects as even the youngest age groups have a significant and negative impact on GPA in ninth grade from severe illness. Lastly, because missing GPAs, due to absence from the exam, are non-randomly distributed between children with and without severe illnesses, issues with selection bias arise. We measure the size of the bias by imputing the missing GPAs with the most extreme cases on the grading scale, in a bounding exercise similar to Genowski et al. (2018). We find that our main results hold, even when we assume the unlikely best-case scenario (highest grade) of the individuals with missing grades.

Not all incidences of diagnoses included in this study are exogenous events. The chances of becoming ill is not the same across social background, and hence social selection might be present due to a problem of omitted variables. The social selection bias occurs, as the incidence of an illness is not always a random event. Parental behavior and capabilities affects the child's risk of becoming ill. We address the selection issue by arguing that several illnesses are in fact exogenous events, and by including a wide range of background characteristics that are correlated with the social selection.

The paper is outlined as follows. First, we briefly go through the relevant literature and describe the Danish context of our analysis. We then present our data and construction of variables. Then we introduce the empirical setup and go through the empirical approach including selection biases. In the next section, the descriptive results are described followed by the estimation results. We finally discuss and conclude on our findings.

Literature review

Several papers document the link between childhood illness and lower school performance. Case et al. (2005) find that having a chronic condition at the age of 7 or 16 lowers the number of passed exams by respectively 15 and 10 per cent. They also find that chronic health conditions in childhood negatively affect the probability of being full-time employed at the age of 33 and 42. Maslow et al. (2011) find similar results for young adults with either cancer, diabetes or epilepsy. They find that these types of chronic illnesses are associated with 10 per cent lower probability of graduating from high school. A Danish study by Andersen et al. (2017) shows that the GPAs of children who have survived cancer are significantly lower than the average of their classmates. This study also finds that school performance differed substantially between cancer types and their results rely on severe types of cancer, e.g. CNS tumors and leukemia. Gurney et al. (2009) also find heterogeneous effects on educational outcomes across different types of cancer. Brain cancer survivors are 20 per cent less likely to attend college than their siblings, whereas for example bone cancer survivors are 2 per cent more likely to attend college.

In line with this study, several other studies demonstrate the importance of considering diversity of severity within diagnoses. For example, half of individuals with a single concussion go on to experience long-term cognitive impairment, while the other half experience significant improvement over the first months following a concussion (McInnes et al., 2017). Epilepsy and diabetes type 1 are also examples of diagnoses for which, some patients are able to uphold a normal life, while other patients are extremely burdened and experience a great deal of hospitalization. The diversity of the severity of illness are also considered by Köhler-Forsberg et al. (2018), who finds a significant association between the number of hospitalizations for infections and lower GPA in a nationwide study of infections among Danish children. Hospitalizations for infections are also associated with a reduced probability of completing ninth grade. Furthermore, their results show that infections within the first 12 months of life and again after age 10 represents the lowest outcomes.

Many studies document that childhood health shocks have long lasting negative consequences on adult outcomes (e.g. Almond et al, 2017; Currie et al, 2008). Also, age-differentiated effect of a health shock on cognitive development are demonstrated by Andersen et al. (2017) who finds lower grades in survivors diagnosed with cancer below the age of 6, but not in those diagnosed in older age groups. In terms of brain unjuries, Prasad et al. (2017) find that students who had suffered a traumatic brain injury at young ages had lower functional academic skill ratings compared to those injured at older ages. Hald Andersen (2017) analyzes how environmental shocks and toxic stress affect the development of a child's developing brain and cognition. She finds that as the brain develops in stages, stress experienced at a given age may have more or less detrimental consequences depending on the specific brain areas under development at the time of the shock. The findings indicate that shocks experienced around age 1-3 and 11-14 reduce the children's performance in math and languages by the end of primary school.

At the same time, there is also evidence of adverse health shocks in childhood resulting in mitigating effects on outcomes in adulthood. Gensowski et al. (2018) exploit quasi-random variation in the poliomyelitis epidemic in Denmark in 1952. For some patients, the infection resulted in a passing period of sickness, while other patients experienced permanent physical disabilities although no impairment of cognitive abilities. This variation caused a shift in the disabled children's comparative advantages, where the permanently disabled children turned out to obtain more education and have higher probability of working in white-collar and computer-demanding jobs.

A major challenge in these types of studies is that the probability of becoming severely ill is not necessarily random. Social background and parental ability are examples of characteristics that could be correlated with the incidence of illness (Danish Health Authority, 2012). However, many papers argue that several chronic diseases and disabilities such as epilepsy, muscular dystrophy, cerebral palsy, missing limbs and cancer are in fact exogenous events in terms of the probability of incidence (e.g. Loft, 2011; Corman and Kaestner, 1992; Joesch and Smith, 1997; Reichman et al., 2004; Holmbroe et al., 2006). This study relies on these findings and assume that these illnesses are in fact exogenous events.

The Danish context

The social and living conditions in Denmark, such as education, income, housing conditions, work environment and the organization of the healthcare system, all play an important role in ensuring a healthy life. In Denmark, all Danish citizens are entitled to publicly financed healthcare. The healthcare system is universal and based on the principles of free and equal access to healthcare for all citizens. The healthcare system is predominantly financed by general taxes (Ministry of Health, 2016). In 2017, total Danish health expenditure amounted to 10.2 per cent of GDP of which 84 per cent is tax financed. For comparison, the OECD average health expenditure as percentage of GDP is 8.9 (OECD, 2018).

In Denmark the healthcare system for children operates at three levels: primary, secondary, and tertiary. Primary healthcare is tax-funded and hence free for all patients. All Danes are entitled to be listed with family doctors who work as gatekeepers to the secondary healthcare at the hospitals (except for emergencies). Tertiary healthcare is known as specialized consultative healthcare usually on referral from primary and secondary healthcare for advanced medical investigation. Private health insurance for children covering accidents and illness, is also available to complement the public system. However, few children are covered by health insurance against illness and the health insurances primarily play a small role regarding compensation and not access to treatment.

Education in Denmark is also provided free of charge. Children must either attend public school or obtain education at an equal level in e.g. a private school or homeschooling. Private schools often charge small tuition fees and the rest of the expenditures are provided by the government. Public expenditure on education and training corresponds to 7 per cent of the Danish GDP and around 13 per cent of total public expenditure (OECDstat, 2018).

For the cohorts considered in this paper, 9 years of education was compulsory for all Danish children. The requirement takes place from first grade, or the calendar year in which the child turns seven, and continues until the ninth grade[5]. However, it is possible to postpone school by one year if the child is not considered to be ready. The child therefore usually turns either 16 or 17 in ninth grade. Some children do not attend the final examination in ninth grade, which is mainly due to learning disabilities, illness or social reasons, and prior to 2007 the final exams were not even mandatory (Danish Economic Councils, 2007).

In this study, we focus on the final examination in ninth grade, where all students are required to complete general tests of academic achievement in different subjects. The mandatory subjects are Danish, Mathematics, English and Natural Science. The exams can be either oral or written, and all written exams are the same across all schools in a given school year. Concerning oral exams, the teacher sets the topics/assignments, hence, they might differ from one class to another. For this reason, this paper uses the grades from written exams only to calculate the GPA. Reporting of exam grades and teacher evaluations is a legal requirement for all schools. The Danish grading scale scores on a 7-point scale ranging from -3 (lowest) to 00, 02 (pass), 4, 7, 10 and 12 (highest). The Danish 7-point scale corresponds to the European Credit Transfer and Accumulation System (ECTS) grading scale, where -3 is F, 00 is Fx, 02 is E, 4 is D, 7 is C, 10 is B and 12 is A.

[5] The vast majority of children in our study also attended one year of optional pre-school (børnehaveklasse), which is compulsory from 2009.

Data and variables

Data and sample

This paper uses individual administrative data from the Danish National Patient Registry (DNPR) and the Nationwide Danish Civil Registration data (NDCR). The two data sources are used to examine the link between childhood illness and the GPA at the end of compulsory school in ninth grade.

The DNPR is collected by the Danish Health Data Authority and has registered all hospitalizations and ambulatory visits since 1977. The data provide a source to identify diseases, time specifications for date and time of admission and discharge from Danish hospitals, contact reason etc. For each patient contact, one primary diagnosis and optional secondary diagnoses are recorded according to the International Classification of Diseases (ICD-10). The DNPR also contains the Medical death statistics, providing information on all deceased persons who lived in Denmark and their cause of death.

The DNPR data are linkable at the patient level with the register data from the NDCR (Schmidt et al., 2015). The register data is collected by Statistics Denmark and covers all Danish residents. The residents each have a unique personal identification number. The data cover many different aspects; educational attainment, demography, social status, family background, income etc. Using the personal identification number, we are able to link registers and identify families.

Our analysis relies on data from 1993-2016, and covers all children born in Denmark in 1994-1999 who lived in Denmark from birth to age 17. We observe the children every year from birth until they finish ninth grade, and record every visit to the hospital due to a somatic illness or accident covered in our groups of diagnosis (see table 1) until the age of 14. Hence, we restrict our sample to children with an illness occurring minimum one year prior to the outcome, because this study focuses on the medium and long run effects of severe illness. We restrict the sample to individuals with non-missing values for key variables.

The long panel and the possibility of linking members of families make our data well suited for studying the correlation between childhood illnesses and school performance, while controlling for socioeconomic background.

Childhood illness

The main explanatory variables of interest in this paper are indicators of whether the child has suffered from a severe illness. However, the term *severe illness* is wide-ranging and must be accompanied by a narrower definition. This section provides such specification.

A list of diagnoses is constructed by using a combination of information from the Danish Patient Associations, the Danish Health Data Authority, medical literature, medical experts and insurance policies for child health insurance in Danish and Swedish insurance companies. The diagnoses included in this study are divided into 46 subgroups, and can be found in appendix 1[6]. The subgroups are aggregated into eight broad diagnosis groups, see table 1. They are created for the purpose of compiling diagnoses with similar features and consequences and to

[6] A complete list of the underlying ICD10 codes can be send upon request.

minimize small sample issues and other requirements related to empirical estimation.

We exclude illnesses with few observations - even if they are consistently covered by child health insurance. We do this to reduce concerns with small sample sizes. The excluded illnesses include mainly heart diseases, diseases of blood and HIV, greatly reduced hearing and blindness.

Concussions are the mildest and most common type of traumatic brain injury, and they are rarely included in other studies that considers a broad range of illnesses. However, a review of the literature on concussions and cognitive impairment (McInnes et al., 2017) finds that, among half of individuals with a single concussion demonstrate long-term cognitive impairment. Therefore, we include the diagnosis, but as a separate group due to high frequency.

Table 1 Severe illnesses included in the study

Diagnosis groups	***Examples of illnesses in the group***
Cancer	*Leukaemia, Brain tumour, Soft tissue Sarcoma*
Muscular dystrophy and arthritis	*Arthritis, Arthrosis, Legg-Calvé-Perthes disease*
Neurological disorders	*Epilepsy, Apoplexy, Stroke, Brain injuries, Benign brain tumour, Sclerosis, Meningitis*
Diseases of the inner organs	*Chron's disease, Ulcerative colitis, Renal insufficiency*
Hormonal and metabolic disorders	*Diabetes type 1, Diabetes type 2, Cystic fibrosis*
Respiratory diseases	*Cronical bronchitis, COPD*
Traumatic injuries	*Fraction and lesions of the spine, spinal cord and nerves, crushing lesions, traumatic amputations*
Concussion	*Concussions*

Notes: Appendix 1 lists the subgroups of the eight diagnosis groups.

The used definition of severe illnesses excludes several diagnoses that are highly prevalent among children, even though they are believed to induce considerable impacts. A large group consist of mental disorders like anxiety, depression, ADHD and schizophrenia. We disregard mental disorders, as they are not easily identified in the data, because the diagnosing is often carried out outside of the hospitals. Another large group excluded from the study is children with asthma. Even though asthma is the leading cause of school absence and hospitalizations of children (U.S. Environmental Protection Agency, 2006), there is no evidence suggesting that asthma in childhood on average has any effect on school achievements (Annett et al., 2000; Gutstadt et al., 1989).

Construction of illness variables

Using the DNPR, we identified the children who have been diagnosed with one of the selected diagnoses. The study considers all incidences among these diagnoses occurring from age 0-14. We construct two types of illness measures based on the diagnoses. The first is an indicator of whether or not a child has been diagnosed with a severe illness included diagnosis list. The second measure distinguishes between the eight diagnosis groups, by constructing an indicator for each group.

By our definition, the course of an illness begins the first time we observe that an individual has visited the hospital with a diagnosis in a subgroup and ends the last time we observe a visit with a diagnosis in the same subgroup. For example, if a child is admitted to the hospital with acute lymphatic leukaemia and later with the same or even a different kind of leukaemia – we record the two visits as the same illness. If the child is admitted to the hospital due to a different kind of cancer, e.g. lung cancer (belonging to another subgroup of cancer) – it is treated as a sequela. By consequence, two concussions are recorded as one, even if they are years apart. A significant obstacle is to measure the severity of an accident or an illness because a clear objective measure is often not available even in administrative data. We will consider the length of hospitalization as a proxy for severity.

Examining data, there is clearly a state dependence in diagnoses. A child diagnosed with cancer has a 25 per cent chance of being diagnosed with cancer again, whereas children with no prior cancer diagnosis have 0.2 per cent chance of getting cancer (according to our findings). The likelihood of sequelae (defined as a second new diagnosis within the same diagnosis group) is highest for children who suffered from diseases of inner organs (30 per cent), second is hormonal and metabolic disorders (21 per cent), and lastly neurological diseases (8 per cent). In our data, we allow children to have sequelae within the same subgroup of diagnosis, however, children diagnosed with two or more non-related illnesses covered in our group of diagnoses, are excluded from the sample.[7] This leads to around 2,500 children being excluded.

We also examine how severity of illnesses affects the results. Severity of an illness is measured as the length of hospitalization. The hospitalization length is constructed as the cumulative measure of the number of days of admission to a hospital due to diagnoses within the same diagnosis subgroups. We construct a variable of hospitalization days within the following ranges: 0 days (only ambulatory visits), 1 day, 2-10 days, 10-100 days and above 100 days. Second, the age at diagnosis is included in the model to investigate age-differential effects of severe illness and identify potential long run effects. Age at diagnosis is defined as the child's age at the first visit to the hospital with a diagnosis included in this study. Specifically, we consider the following age intervals: 0-2, 3-5, 6-10 and 11-14, corresponding to important stages of childhood: early childhood, preschool years, early compulsory school, and late compulsory school, respectively.

School performance

In this paper we wish to examine an impact on children's cognitive abilities measured as school performance. For this purpose, we use grades from the end of compulsory school in ninth grade. From the administrative data, we use the registry of compulsory school grades compiled by the Ministry of Education based on school reports. The register holds both exam grades and continuous assessment grades (given at least twice during a school year).

We construct the GPA based on grades from the written exams in Danish (reading test, written presentation and dictation) and Mathematics (arithmetic without a calculator and arithmetical problem solving). First, we construct a GPA for each subject, and then a final average based on the two subject-specific GPA.

Exam grades are often considered the best measure of academic performance, as both the teacher and an external examiner assess performances, making teacher-

[7] Diagnoses are unrelated if they appear in two different subgroups of diagnosis. The only exception is children with a concussion who are subsequently diagnosed with a neurological disease.

specific bias less of an issue (Fallesen, 2015). In particular, written exams are argued to have less teacher-specific bias, as they are identical across schools[8] and therefore easy to compare (Nielsen and Rangvid, 2012). However, not all children attend the exams, and it seems fair to assume that it is the weaker children with less support from home or suffering from a severe illness, who are likely to stay away from an exam.

About 10 per cent of the children in our sample do not attend the final exams in ninth grade, and by excluding these children, we will most likely underestimate the impact of childhood illness on GPA. To support this, we also find that the average assessment grades are in fact lower for those not attending the exams compared to those attending. We therefore replace the missing exam grades with the continuous assessment grades when possible, which successfully add grades to 35 per cent of the non-attenders. The potential bias resulting from the remainder on the non-attenders will be addressed later in the section of the empirical approach.

Background characteristics

On the assumption that parental capabilities and children's health and cognitive endowments interact with both illness and school performance, we include a long list of background characteristics of the child and its parents. The background information is taken from the administrative data, NDCR. We include gender, ethnicity, number in birth order, year of birth, family type at birth (living with both parents, either of the parents or none of the parents) and birth weight in child characteristics. The latter is included to control for the child's health endowment at birth. Among parent characteristics, we include mother's age at birth, if father is unknown, if at least one parent is deceased, parents' highest educational level and parent's socioeconomic status. Appendix 2 lists all the explanatory variables included in the study and presents means for all children with severe illness and for children without. We will discuss the variables in the descriptive results.

Empirical approach

Baseline models

In order to examine whether severe childhood illness affects children's school performance, we consider the following regression setup. The simplest version of our model, which we refer to as the baseline, is:

$$GPA_i = \alpha + \beta I_i + X_i'\gamma + \varepsilon_i \qquad (1.1)$$

where GPA_i denotes the child's GPA when completing compulsory school in ninth grade. X_i is a vector of controls (including child and parental characteristics and cohort dummies). I_i is an indicator variable denoting whether the child has suffered from one of the severe illnesses included in the eight diagnosis groups, see table 1. The coefficient β is of primary interest. It captures the average difference in GPA of children with and without a severe illness.

As the eight diagnosis groups represent a heterogeneous group of illnesses and we expect them to affect children in various ways, we will also consider a more detailed version of the baseline model. In model 1.1 we replace the illness

[8] The written exams in compulsory school are produced by the Ministry of Education.

dummy, I_i, with a vector of dummies representing each of the eight diagnosis groups, D_i':

$$GPA_i = \alpha + D_i'\beta + X_i'\gamma + \varepsilon_i \qquad (1.2)$$

Our outcome variable is only observed once, hence the cross sectional data setup. All time varying variables are observed in the same year as the child's birth[9], to ensure that the parental characteristics are not affected by the child's course of illness.

Extensions

To understand the impact of the level of severity associated with illness and long run effects of illness on children's cognitive abilities, we introduce interaction terms in our baseline models. This allows us to study the variation in severity (measured by length of hospitalization) and variation in the time span between time of diagnosis and outcome (measured by age at diagnosis). By extending the baseline models, we examine both the interactions across diagnosis groups and within. The reference group for all interactions are children who never experienced a severe illness.

Length of hospitalization

Hospitalization is a direct measure of both the severity of illness and absence from school. Given that health capital (low level of severity) and the amount of time children spend in the classroom are positive predictors of children's school performance, hospitalization is expected to be negatively correlated with school performance.[10]

We include the length of hospitalization as five interval dummies to interact with the illness status. As such, we explore whether the length of hospitalization during the course of a severe illness is associated with lower GPA. We estimate mean differences in GPA according to the number of days the child is hospitalized using the following linear regression setup:

$$GPA_i = \alpha + (I_i \times hospitalization_i')\beta + X_i'\gamma + \varepsilon_i \qquad (2)$$

Age at diagnosis

To investigate the composite effects from severe illness and time of diagnosis on school performance, we introduce variation in the time span between the school exam and the time of diagnosis, by including age at diagnosis in our model. Age-differentiated effects are included in the baseline model by interacting illness status with intervals of age at the time the child was diagnosed. The following model is estimated:

$$GPA_i = \alpha + \times (I_i' \times age\_at\_diagnosis_i')\beta + X_i'\gamma + \varepsilon_i \qquad (3)$$

[9] Parental socioeconomic status is observed the year before the child's birth, to ensure maternal leave does not influence the socioeconomic status.

[10] We want include both effects from the length of hospitalization, because they are all a direct consequence of the illness.

In essence, we compare the effect of illness status on GPA in the long run (age at diagnosis 0-2) to the medium-run (age 3-5 and 6-10) and short run (age 11-14).

Identification

Identification arises when all factors of being selected into illness have been considered in the models. Hence, our research design relies on the assumption that the incidence of being diagnosed with an illness is uncorrelated with unobservable determinants that also impact individual school performance. However, because available data are often deficient in practice, selection issues may arise.

Social selection

Selection in the incidence of illness implies that both the likelihood of being exposed to a health shock and children's subsequent school performance are simultaneously affected by third party variables – e.g. parental education and other indicators of parental capital (see e.g. Wilkinson, 1986). If it is not possible to include all such determinants, we potentially face an omitted variable bias. Unobserved characteristics such as choice of lifestyle (smoking, eating healthy, exercise etc.), level of parenting skills and cognitive abilities, and child's cognitive ability prior to the health shock, may be correlated with both the incidence of illness and GPA, and thus bias the results. For example, parents with lower capabilities are more likely to have children with lower cognitive endowment (Bernal and Keane, 2008) but are also more likely to have children who are exposed to injuries (Laursen et al., 2006). This kind of selection bias is often referred to as social selection, and we will do the same.

While we cannot control directly for all such factors, due to data constraints, we can control for other factors likely to be correlated with the omitted variables. For instance parental education and socioeconomic status, which are commonly used as proxy for parental skills and hence the cognitive endowment of the child (Erola et al., 2016). In addition, birth weight is a very good predictor of the overall health endowment of a child and is highly correlated with the mother's pregnancy behavior (Currie et al., 2008; Gupta and Simonsen, 2007). For example, smoking during pregnancy is linked to higher risk of respiratory diseases (Hermann et al., 2006). Both parental background and birth weight are widely applied in similar models of related literature (Almond et al., 2017; Currie et al., 2008; Case and Paxson, 2010; Gupta and Simonsen, 2007; Fallesen and Bernardi, 2018).

There is reason to expect that social selection is an issue for some of the diagnosis groups included in this study. However, selection into illness is not directly a self-selection problem since no child actively chooses to become ill and, likewise, parents do not usually intend for their child to become ill. Several studies argue that a number of the diagnoses included in this study are in fact exogenous events. For instance, chronic illnesses and disabilities such as epilepsy, muscular dystrophy, cerebral palsy, missing limbs and cancer (e.g. Loft, 2011; Corman and Kaestner, 1992; Joesch and Smith, 1997; Reichman et al., 2004; Holmbroe et al., 2006). Given that this holds in our setting, it would correspond to these four diagnosis groups being exogenous: cancer, muscular dystrophy and arthritis, diseases of inner organs and hormonal and metabolic diseases. When comparing baseline results with an unadjusted model without background characteristics, we partly address the question of whether the diagnosis groups are in fact exogenous.

There is also another social gradient related to illness due to social differences in the impact of an illness, which occurs *after* the incidence of the health shock (Danish Health Authority, 2012). Several studies find that parents with higher socio-economic status have a higher tendency of earlier diagnosing (Moth et al., 2008), higher health literacy (Dewalt et al., 2007), and lower mortality (Syse et al. (2012), suggesting that children of high-ability parents are less affected by illness compared to children with low-ability parents (suffering from same illness). This study does not distinguish between the direct impact on GPA set in motion by the incidence of illness alone and the indirect impact that is due to social difference in parents response to their child's illness, as both impacts occurs *after* the incidence of illness. However, bias might occur in cases, where hospital admission happens long after the illness is detected, because we do not observe the incidence of illness before the child is admitted to the hospital.

Selection bias due to non-exam attendance

Unfortunately, even if a health shock is a random event, one cannot use simple treatment-control differences to estimate the effect on grades. This is because the GPAs are only observed for children who attend the exam. This gives rise to the classical sample selection problem, which is pervasive in applied econometrics (see Heckman, 1979).

This kind of selection bias arises because missing GPAs are non-randomly distributed between children with and without severe illnesses. Because children with a severe illness are less likely to attend the exam, grades might be correlated with the likelihood of exam attendance, and the effects of a health shock on GPA are therefore underestimated. This is inferred from the assumption that children with an illness who are highly affected in their daily lives are less likely to perform well even if they had attended the exam and are at the same time less likely to attend the exam in the first place.

The selection issue is primarily due to exam non-attendance and deaths prior to ninth grade. To address the sample selection problem, we first impute the missing exam grades by using the continuous assessment grades when possible. This solves the problem for about a fourth of the students with missing exam grades. On the imputed data we perform a bounding exercise analogues to Lee (2009). Because the support of our outcome variable is bounded, we can construct "extreme-case scenario" bounds of the health shock, by assuming that the missing data on grades, due to non-attendance, are either at the top (grade = 12) or at the bottom (grade = -3) of the distribution. This allows us to compute the largest and smallest possible effects on grades consistent with the data observed. This procedure does not solve the selection bias due to non-exam attenders but indicates the maximum magnitude of the problem.

Descriptive results

We identify 381,117 children born between 1994 and 1999, corresponding to approximately 63,500 children in each birth cohort on average. 340,005 children completed the exam in ninth grade during our study period, and 14,539 children where imputed with their continuous assessment grades, leading to 354,544 children with a GPA. 34,337 children had been diagnosed with a severe illness included in the eight diagnosis groups, corresponding to 9 per cent of the population.

Appendix 2 lists all the explanatory variables used in the analysis and presents means for children with and without an illness. Boys, second born children, and children with low birth weight seem to be over-represented among children with an illness. Likewise, children with an illness are more frequent in one-parent families, or in families with parents having limited education, being welfare claimants or being unemployed. This indicates that social selection is present in our study, hence including the background characteristics in our model is essential.

Table 2 shows frequencies and school performance for all children and for children who suffered from severe illness for each diagnosis group. The first column of table 2 reports the number of children in our sample while the second column report their average GPA. The remaining columns reports exam attendance rate, share of students with non-missing GPA (after imputation with continuous assessment grades) and their mortality rates.

Concussions are the most prevalent illness in our sample, next is neurological diseases – mostly due to epilepsy. Children who suffered from either cancer or neurological diseases are the least likely to attend exams and therefor have missing grades. Children with neurological diseases are also in the group of children with the lowest GPAs together with children suffering from traumatic injuries and respiratory diseases. We also see that the exam attendance rate and the share of non-missing GPAs are decreasing substantially as the length of hospitalization increases.

Table 2 Children with an illness – frequency, exam attendance and GPA

Description	Number of children	Grade point average	Exam attendance rate	Non-missing GPA rate	Mortality rate
All children	381,117	6.5	89.2	93.0	0.3
Children without illness	346,780	6.5	90.0	93.6	0.2
Children with illness	34,337	6.1	82.0	86.8	1.3
Cancer	703	6.3	68.6	73.3	19.5
Muscular dystrophy and arthritis	1,863	6.4	88.3	92.6	0.2
Neurological diseases	6,802	5.9	64.2	70.4	3.3
Inner organs	597	6.3	84.8	88.6	1.5
Hormonal and Metabolic	1,474	6.4	84.4	89.6	0.7
Respiratory diseases	2,045	6.0	84.1	88.9	0.3
Traumatic injuries	2,503	5.9	83.2	88.3	1.3
Concussion	18,350	6.3	87.8	92.2	0.2
Hospitalization					
0 days	16,067	6.2	85.6	90.1	0.4
1 day	10,471	6.2	84.8	89.5	0.7
2-10 days	4,989	6.1	76.8	82.5	1.5
10-100 days	2,542	5.9	62.0	67.5	6.6
over 100 days	268	6.2	44.0	47.8	27.6

Notes: Non-missing GPA also includes those who did not attend the exams and had continuous assessment grades.

As discussed, the different diagnosis groups contain a great deal of diversity, and table 3 explores how much hospitalization varies within the groups as we assume high correlation between severity of an illness and length of hospitalization. Table

3 also presents mean and standard deviation of age at diagnosis for each diagnosis group.

The length of hospitalization varies greatly across and within the eight diagnosis groups. The largest variation is found in the length of hospitalization for cancer patients. Also, hospitalizations due to neurological diseases, diseases of inner organs, and hormonal and metabolic disorders vary greatly in length. Hence, there is scope for examining variation in the severity within diagnosis groups by measuring how the estimated effect of the baseline model is affected, when taking account for length of hospitalization.

Table 3 Illness and age at diagnosis and hospitalization

Description	Days of hospitalizations		Age at diagnosis	
	--Mean--	-Std. dev.-	--Mean--	-Std. dev.-
Illness – all	4.4	17.5	5.9	4.6
Cancer	69.7	65.2	5.7	4.6
Muscular dystrophy and arthritis	3.2	8.5	7.7	4.3
Neurological diseases	7.7	20.0	5.6	4.4
Inner organs	12.4	24.7	10.5	3.8
Hormonal and Metabolic	12.8	21.0	8.4	4.1
Respiratory diseases	1.9	5.5	1.4	2.1
Traumatic injuries	1.9	6.0	8.2	4.8
Concussion	0.6	0.9	5.6	4.4

Likewise, there is great variation in age at diagnosis across diagnosis groups, while the variation within is around 4-4.5 for most of the diagnosis groups – only respiratory diseases seem to be diagnosed very early in childhood for the vast majority of patients. As age only takes on values ranging from 0-14, a standard deviation of 4.5 is significant.

Estimations results

This section quantifies the effect of severe illness on the GPA of ninth graders using the empirical approach previously outlined.

Main results

Table 4 presents our main results when estimating the effect on GPA of suffering from severe illness. In model 1.1 we measure the effect of illness in both an unadjusted (no controls) and an adjusted model (including background characteristics). The results from the unadjusted model suggest that severe illness in childhood lowers GPA by 0.36 grade points on average (mean is 6.5). When covariates are added to the model, the effect of severe illness decreases to 0.26, corresponding to a 4 per cent reduction in GPA compared to children without an illness. These findings suggest that this significant reduction in GPA cannot solely be attributed to differences in terms of birth weight, family situation, parental socioeconomic and academic resources and demographic characteristics, measured at the point of childbirth.

Model 1.2 includes a dummy for each diagnosis group, and the results indicate a large variation in the effect on GPA across diagnosis groups, but also between the unadjusted and adjusted model. In the adjusted model, the biggest significant effect is estimated for neurological disorders with a reduction of 0.59 in GPA, whereas the smallest significant effect is found for concussions with 0.18 in reduction. The other diagnosis groups lowers GPA by 0.45 on average for traumatic injuries, 0.24 for inner organs, 0.22 for cancer, 0.21 for respiratory diseases. The effect on GPA from hormonal and metabolic diseases and muscular dystrophy and arthritis are not significant.

As expected, we find that not all illnesses are randomly distributed, and are therefore not an exogenous shock. This is true for respiratory diseases, traumatic injuries, and concussions, for which the coefficient estimates falls with 30-60 per cent between the unadjusted and the adjusted model (differences in coefficients estimates are presented in the last column of table 4). For instance, the risk of developing a respiratory disease is strongly correlated with parents' smoking behavior (for example Lanari et al., 2015; Lannerö et al., 2006; Hermann et al., 2006), so when controlling for parental education and socioeconomic status, we expect the estimates to drop. Likewise, for conditions associated with injuries, such as traumatic injuries and concussions. Previous research find that social background is correlated with the probability of children being injured (Laursen et al., 2006), and parents with higher education might be less likely to leave their children unattended and pass on better judgements skills, which might explain why these diagnosis groups are non-randomly distributed.

Table 4 baseline model

Dependent variable: GPA	Model 1.1		Model 1.2		
	Unadjusted	Adjusted	Unadjusted	Adjusted	Diff. in coefficients
Illness dummy	-0.364***	-0.257***			
Cancer			-0.205*	-0.216**	0.011
Muscular dystrophy and arthritis			-0.094	-0.079	-0.015
Neurological disorder			-0.680***	-0.585***	-0.095***
Inner organs			-0.273**	-0.237**	-0.036
Hormonal and Metabolic			-0.097	-0.103	0.006
Respiratory diseases			-0.576***	-0.212***	-0.364***
Traumatic injuries			-0.685***	-0.449***	-0.236***
Concussions			-0.266***	-0.177***	-0.089***
Controls	No	Yes	No	Yes	
Mean, dept. var.	6.5126	6.5126	6.5126	6.5126	
R-squared (adjusted)	0.0014	0.2150	0.0019	0.2152	
Observations	354,544	354,544	354,544	354,544	

* Significant at 10%. ** Significant at 5%. *** Significant at 1%.

Notes: All regressions are estimated using OLS with heteroskedastic robust standard errors. In the regressions with controls, we control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, either one or none), parental educational level and socioeconomic status prior to birth. All estimates of the adjusted models are presented in appendix 3.

The last column of table 4 show that the estimated coefficients does not change significantly for cancer, muscular dystrophy and arthritis, diseases of inner organs and hormonal and metabolic diseases between the adjusted and the unadjusted model. This indicates that a health shock caused by any diagnosis included in the eight diagnosis groups is uncorrelated with the observed background characteristics, supporting the findings of previous literature, that these health shocks are exogenous events (e.g. Loft, 2011; Corman and Kaestner, 1992; Joesch and Smith, 1997; Reichman et al., 2004).

The social selection issue relates to the likelihood of becoming ill in the first place, but there also exists social differences in the impact of the illness that arises after the illness has occurred (e.g. parental response). The latter might still influence our results, as we do not necessarily observe the actual point in time where the illness is detected, but first when the child enters the hospital sector. If these social differences in the impact are affecting our results, they should be picked up in the differences between the estimated coefficients of the unadjusted and the adjusted model for the four diagnosis groups considered to be exogenous shocks. However, as the differences are insignificant for the four diagnosis groups in question, this bias is most likely not present. This could also suggest that the bias is not be present for the endogenous diagnosis groups either, but we cannot identify this directly.

Lastly, a possible issue might arise when background characteristics are measured around childbirth. At the time of childbirth, parental capital could be difficult to observe because many parents has not yet finished their education. To address this issue, we have estimated both of our baseline models considering only the children being diagnosed at age 11-14. In essence, we estimate a model where time-varying background characteristics are observed around childbirth and one where they are observed at age 10. The two versions delivers very similar estimates of the coefficients of both the overall illness status and the eight diagnosis groups.[11]

Length of hospitalization

Above we implicitly assume that the effects of having an illness are equal for everyone in the same diagnosis group. However, there are reasons to believe that this is not the case, and studying the data in DNPR shows that the amount of time a child is hospitalized varies greatly within each diagnosis group (see table 3).

Given that health capital (low severity) and the amount of time students spend in the classroom are positive predictors of student achievement, absence due to hospitalization will be negatively correlated with GPA. We include five hospitalization dummies to compare the effect on GPA for children who have been hospitalized 0 days (only ambulatory visits), 1 day, 2 to 10 days, 10 to 100 days and over 100 days, respectively, to children without an illness (the reference group).

[11] The results are available upon request.

Table 5 Illness and length of hospitalization

Dependent variable: GPA	Model 2	
	GPA	SE
Illness x 0 days	-0.230***	[0.020]
Illness x 1 day	-0.224***	[0.024]
Illness x 2 to 10 days	-0.330***	[0.038]
Illness x 10 to 100 days	-0.477***	[0.059]
Illness x over 100 days	-0.553**	[0.215]
Controls	Yes	
Mean, dept. var.	6.513	
R-squared (adjusted)	0.215	
Observations	354,544	

* Significant at 10%. ** Significant at 5%. ***Significant at 1%.

Notes: The regression is estimated using OLS with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, either one or none), parental educational level and socio-economic status prior to birth.

The results of table 5 show that being hospitalized with severe illness for a longer period of time is associated with a lower GPA in ninth grade. The effect is substantially larger for illnesses that leads to long hospitalization. Standard errors indicate that there is no significant difference in grades for children hospitalized 1 day compared to ill children with only ambulatory visits (0 days). However, the average GPA decreases significantly for children who were hospitalized for 2 days or more, and we find the reduction in GPA is highest among children who were hospitalized for 10 days or more.

To examine how severity of an illness is associated with lower grades across each diagnosis group, we consider the detailed version of the baseline model (see appendix 4). The effect of illness status on GPA is increasing with the number of overnight stays at the hospital for all eight diagnosis groups, however not significantly so for cancer, muscular dystrophy and arthritis. For cancer, only children with more than 100 overnight stays at the hospital are significantly affected by their illness status, while muscular dystrophy and arthritis becomes hampering for children's development of cognitive skills when hospitalization exceeds 1 day.

Age at diagnosis

Allowing for age-differential effects of childhood illness on GPA enables us to assess whether the effects from illness status last in the long run. We include four age dummies interacted with illness status in the baseline model. This allows us to examine the effect on GPA for children diagnosed with an illness early in childhood and the effect on GPA for children diagnosed with an illness later in childhood, separately.

Table 6 Age and illness

Dependent variable: GPA	Model 3	
	GPA	SE
Illness x age at diagnoses 0-2	-0.208***	[0.025]
Illness x age at diagnoses 3-5	-0.220***	[0.030]
Illness x age at diagnoses 6-10	-0.309***	[0.029]
Illness x age at diagnoses 11-14	-0.302***	[0.029]
Controls	Yes	
Mean, dept. var.	6.513	
R-squared (adjusted)	0.215	
Observations	354,544	

* Significant at 10%. ** Significant at 5%. ***Significant at 1%.

Notes: The regression is estimated using OLS with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, either one or none), parental educational level and socioeconomic status prior to birth.

The results in table 6 confirms that the results found in the baseline model applies regardless of the age of diagnosis. The estimates for all age groups are significantly negative, and the effect of illness is increasing with the age at which the child is diagnosed. The standard errors (column 2) indicate that the increasing effect is significant. Overall, the results indicate that older age groups are significantly worse off compared to the younger age group (0-2 years), but that severe illness *even* in early childhood has significant effects on future school performance.

We have also estimated the age-differential effects for each diagnosis group and the results are presented in appendix 5. The results indicate that the aggregated version presented in table 6 are highly affected by composition effects from aggregating the heterogeneous diagnosis groups contained in the illness dummy.

Composition effects arise because the overall result are highly affected by prevalence and distribution of the underlying diagnosis groups. For example, some diagnosis groups are primarily diagnosed early in childhood (respiratory diseases), while others are primarily diagnosed later in childhood (diseases of inner organs). Further, because neurological diseases and concussions accounts for 73 per cent of all incidences of severe illness, the significant negative effect on GPA from these diagnosis groups will be highly reflected in the overall result and most likely dominate the effects from less prevalent diagnosis groups (e.g. cancer). Lastly, some diseases might be detected early because of severity, which also might affect the causal interpretation of age. Composition effects also play a role within diagnosis groups – especially for cancer, which is a group with a high variation in levels of severity. It seems however reasonable to assume that concussions might be less affected by composition effects due to prevalence, as concussions are highly frequent in all age groups.

Many of the mean differences reported in appendix 5 are not statistically different from zero, but all significant results are of the anticipated sign. Neurological disorders and concussions seem to cause lower grades across all age groups, while cancer and hormonal and metabolic diseases only seem to have a significant effect if diagnosed early in childhood. The reverse is true for diseases of inner organs and muscular dystrophy and arthritis which only seem to have an effect if diagnosed later in childhood. Respiratory diseases stand out by having a U-shaped effect on the grade average, having a modest but significant negative effect early in childhood, insignificant in the pre- and early school years, and then a significant and substantially negative effect on GPA in the later school years. However, this might be due to a timing issue, as 86 per cent of the incidences are among children aged 0-2.

Whether this result is due to the hypothesis that because health human capital is complementary to skills and "skills begets skill", so that children who suffer early in childhood fall behind and never catch up, or if the lower grade averages are due to effects on future health outcomes, we cannot conclude from this study. However, we can confirm that even in the long run, the effect of childhood illness on children's cognitive abilities seems to persist.

Selection bias due to missing GPAs

All of the results presented above are based on the sample of children attending ninth grade exam and thus obtaining a GPA, and the10.000 children for whom we successfully imputed the missing GPAs with their ninth grade continuous assessment grades. However, around 30.000 children, or 8 per cent of each birth cohort, neither attended the exam or had any assessment grades from ninth grade. This section seeks to quantify the bias associated with leaving out children with missing GPAs from the study.

Table 2 in the previous section reports attendance rates (the proportion of the children who attended ninth grade exam or make-up exams) for all children and for children with an illness. The results indicate that the overall mean masks heterogeneity in attendance rates between children with and without illness and also between diagnosis groups. Children with an illness are more likely to miss an exam, and hence, less likely to be included in our estimation models, which creates a selection bias. Thus, the incidence of illness is negatively correlated with attendance propensities for all eight diagnosis groups, and being diagnosed affects whether children have missing exam grades. This is also evident from the attendance rates in table 2. Even though a proper random health shock implies that children with an illness are comparable to children without an illness at the baseline, they may well be systematically different conditional on attending the exam subsequent to the health shock. As a result, the difference in GPA, as presented in table 4, may be biased and therefore not represent the true causal effect of the health shock.

Table 7 shows results from a regression of exam attendance on illness status and background characteristics. The logit coefficients represents the estimated effects of each diagnosis group on the odds ratio for the dependent variable. The odds ratio is the likelihood of a child with an illness to attend the exam compared to the likelihood of a child without illness to attend the exam. The results in table 7 show that the odds ratio for children diagnosed with a severe illness, in either of the eight diagnosis groups, is well below 1 (representing equal likelihood for exam attendance). This means, that children with an illness are significantly less likely to attend the ninth grade exam compared to children without an illness. This is especially true for children with cancer and neurological disorders, who are 4-5 times less likely to attend the exam compared to children without illness.

Table 7 Predictors of Exam attendance

Dependent variable: Exam attendance	Odds Ratio
Cancer	0.258***
Muscular dystrophy and arthritis	0.872*
Neurological disorder	0.208***
Inner organs	0.619***
Hormonal and Metabolic	0.605***
Respiratory diseases	0.746***
Traumatic injuries	0.638***
Concussion	0.880***
Controls	Yes
Mean, dept. var.	0.91
R-squared (adjusted)	0.215
Observations	354,544

* Significant at 10%. ** Significant at 5%. ***Significant at 1%.

Notes: The regression is estimated using logistic estimation with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, either one or none), parental educational level and socioeconomic status prior to birth.

We address the issue of selection bias due to missing GPAs by implementing the bounding exercise described in the empirical approach. We include all children that leave our data prior to the exam at the end of ninth grade, with either their most favorable outcome (grade = 12) or least favorable outcome (grade = -3), and perform our 3 models on the resulting sample with imputed variables. Table 8 presents the results from this bounding exercise for outcomes in our baseline models 1.1 and 1.2, model 2, interacting hospitalization and illness, and finally model 3, interacting age at diagnosis and illness.

Table 8 Measuring selection bias

Dependent variable: GPA	Baseline result	Imputed values for missing grades	
		Least favourable (grade = -3)	Most favourable (grade = 12)
Model 1.1 - baseline model			
Illness	-0.257***	-0.751***	-0.127***
Controls	Yes	Yes	Yes
Mean, dept. var.	6.513	5.849	6.895
R-squared (adjusted)	0.215	0.196	0.107
Observations	354,544	381,117	381,117
Model 1.2 - baseline model			
Cancer	-0.216**	-1.844***	0.900**
Dystrophy and arthritis	-0.079	-0.129*	-0.015
Neurological disorder	-0.585***	-2.402***	-0.860***
Inner organs	-0.237**	-0.651***	0.066
Hormonal and Metabolic	-0.103	-0.415***	-0.155**
Respiratory diseases	-0.212***	-0.438***	-0.004
Traumatic injuries	-0.449***	-0.754***	-0.133***
Concussion	-0.177***	-0.228***	-0.110***
Controls	Yes	Yes	Yes
Mean, dept. var.	6.513	5.849	6.895
R-squared (adjusted)	0.2153	0.201	0.109
Observations	354,544	381,117	381,117
Model 2 - hospitalization and illness			
Illness x 0 days	-0.230***	-0.452***	-0.037
Illness x 1 day	-0.224***	-0.487***	-0.003
Illness x 2-10 days	-0.330***	-1.161***	0.324***
Illness x 10-100 days	-0.477***	-2.577***	1.100***
Illness x over 100 days	-0.553**	-4.164***	2.217***
Controls	Yes	Yes	Yes
Mean, dept. var.	6.513	5.849	6.895
R-squared (adjusted)	0.215	0.199	0.108
Observations	354,544	381,117	381,117
Model 3 - age and illness			
Illness x age at diagnoses 0-2	-0.208***	-0.841***	-0.261***
Illness x age at diagnoses 3-5	-0.220***	-0.753***	-0.190***
Illness x age at diagnoses 6-10	-0.309***	-0.792***	-0.073**
Illness x age at diagnoses 11-16	-0.302***	-0.576***	-0.064*
Controls	Yes	Yes	Yes
Mean, dept. var.	6.513	5.849	6.895
R-squared (adjusted)	0.215	0.196	0.107
Observations	354,544	381,117	381,117

* Significant at 10%. ** Significant at 5%. ***Significant at 1%.

Notes: All regressions are estimated using OLS with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, either one or none), parental educational level and socioeconomic status prior to birth.

The results from the baseline model are robust to the bounds, meaning that the negative impact of a health shock persists even in the unlikely case, that all individuals who did not attend ninth grade exam, would have had the most favorable outcome. At the diagnosis group level, only cancer and diseases of inner organs are not robust to the bounds (i.e. the effect is insignificant). Cancer represents a special case, being the diagnosis group with the widest gap between the least- and most favorable outcome. The reason for this wide interval is that more than 30 per cent of children with a cancer diagnosis did not attend the exam in ninth grade.

When estimating models 2 and 3, the results show that whereas the size of the bias is the same size across age groups, the bias increases with the number of days a child is hospitalized. Again, the reason for the widening interval is that the share of children with a missing GPA is strongly correlated with the number of days spend in the hospital, and results in the large bias we see in the effect from hospitalization.

If health capital and time spent in the class room are positive predictors of student achievement, and the length of hospitalization predicts severity, there is reason to believe that the selection causes an upward bias in our results. Children with an illness and a high rate of severity of the illness are under-represented in our outcome variable, and we end up regressing school performance on illness status for children who are well enough to attend the exam in the first place. Hence, we expect the estimated effects to be underestimated in relation to this specific selection issue of exam attendance.

If we assume that children with an illness who are highly affected in their daily lives are less likely to perform well even if they had attended the exam and at the same time are less likely to attend the exam in the first place, the effect of a health shock on GPA are underestimated. We therefore expect the effect to lie somewhere between the estimates from the baseline model and the lower bound.

It is important to state that this exercise simply addresses the bias caused by missing GPAs and does not solved the problem of biased estimates. The results however gives us a good idea of the size and direction of the bias and is good indication of where we should draw our attention for future refinement of the model. One possible improvement is to employ the Heckman two-step procedure.

Conclusion

This paper analyzes the link between severe childhood illness and the GPA by the end of compulsory school in ninth grade in a Danish context. Large numbers of children suffers from a severe illness in childhood (9 per cent according to our findings). Previous research across countries show that there is a strong link between the incidence of being diagnosed with severe illness and socioeconomic outcomes, such as cognitive abilities, socioeconomic status and parental relationship (Andersen et al. (2017); Loft (2011); Currie et al. (2008); Case et al., 2005). However, how severe illness in a broad sense affects school outcomes in a context where all children are entitled to free healthcare have not yet been analyzed. We also contribute with new evidence of the effects on school performance from severe illness as a broad definition and make comparisons across the different diagnosis groups. This is possible because we have access to rich administrative data on all hospitalizations, diagnoses and information on socioeconomic variables of all family members for all Danish children.

We leverage multiple sources of linked data and compare the outcomes of children who were diagnosed with an illness covered by this study (cancer, muscular dystrophy and arthritis, neurological disorders, diseases in the inner organs, hormonal and metabolic disorders, respiratory diseases, traumatic injuries and concussions) and children without illness at ages 0 to 14 born between 1994 and 1999. The strengths of our study include: complete coverage of the population from birth to follow up, a broad definition of childhood illness that include most incidences of severe illness for children, and the use of objective health measures. Access to comprehensive administrative data enables us to control statistically for family characteristics related to the incidence of childhood illness.

Our paper offers several conclusions. The results from our baseline model show that children who suffered from severe illness during childhood perform less well relative to other children. Our findings indicate an overall 0.3 grade point reduction of GPA (4 percent) for children with severe illness, but the overall effect is highly affected by the included diagnosis groups. The largest coefficient estimates are found for neurological disorders (-0.6) and traumatic injuries (-0.5), while cancer, diseases of inner organs, respiratory diseases and concussions have smaller coefficient estimates (around -0.2). The incidence of muscular dystrophy and arthritis and hormonal and metabolic disorders provides insignificant results, but by interacting with the length of hospitalization these diagnosis groups are significant and negative for children hospitalized more than 2 days. These results demonstrate that the effects of severe illness varies greatly both between and within diagnosis groups.

Our findings also suggest that neurological disorders, respiratory diseases, traumatic injuries and concussions, are likely subject to social selection, even after controlling for parental education and socioeconomic background etc. However, the coefficient estimates for neurological diseases and traumatic injuries are large and highly significant, which indicates that not all are explained by social factors.

We measure the severity of an illness as the amount of time spend in the hospital and our results indicate that long hospitalizations increase the adverse effects of severe illness on GPA. Results from the bounding exercise show that the added effect from long hospitalization is underestimated because only 60 per cent of children with over 10 days at the hospital actually attend the exam. If we believe that children who are absent are less able to perform in the first place, the effect of long hospitalization (10 days or above) on GPA is somewhere between the baseline results of 0.5 grade reduction and the lower bound results of 2.6 grade point reduction in GPA.

The aim of this study is to analyse the medium to long run effects of childhood illness, why we constrain our sample to children diagnosed with a severe illness prior to the their 15th birthday. Furthermore, we test if there appear to be long run effects by allowing age at diagnosis to interact with the incidence of severe illness. We find evidence of an adverse effect of severe illness on ninth grade GPA - even for children diagnosed early in childhood.

The findings demonstrate how severe illness affects the child's ability to invest in their human capital. They tend to perform significantly worse in the final exams and they are 50 percent less likely to even attend the exams compare to children without a severe illness. The lower performance in compulsory school might be difficult to catch up or compensate for; hence it is reasonable to fear that impact could be lasting all the way into adulthood. It would be very interesting to extend the study to include young adult and adult outcome for future work.

References

Andersen, K.K., A.K. Duun-Henriksen, M.H. Frederiksen and J.F. Winter (2017). Ninth grade school performance in Danish childhood cancer survivors. *British Journal of Cancer, 116*(3), 398-404.

Almond, D., J. Currie and V. Duque (2017). Childhood Circumstances and Adult Outcomes: Act II. *NBER Working Paper Series,* 23017.

Annett, R., E. Aylward, J. Lapidus, B. Bender and T. Duhamel (2000). Neurocognitive functioning in children with mild and moderate asthma in the childhood asthma management program. The Childhood Asthma Management Program (CAMP) Research Group. *The Journal of Allergy and Clinical Immunology, 105*(4), 717-24.

Bernal, R. and M. Keane (2008). Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers'. *International Economic Review* 49, 1173-1209.

Case, A., A. Fertig, and C. Paxson (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics, 24*(2), 365-389.

Case, A. and C. Paxson (2010). Causes and consequences of early-life health. *NBER Working Paper Series,* 15637.

Corman, H. and R. Kaestner (1992). The effects of child health on marital status and family structure. *Demography, 29*(3), 389-408.

Cunha, F. and J.J. Heckman (2007). The Technology of Skill Formation. *NBER Working Paper Series,* 12840.

Currie, J., M. Stabile and L. Roos (2008). Child Health and Young Adult Outcomes. *NBER Working Paper Series,* 14482.

Danish Economic Councils (2007). Danish Economy, fall 2007.

DeWalt, D. A., M. H. Dilling, M. P. Pignone and M. S. Rosenthal (2007). Low Parental Literacy Is Associated With Worse Asthma Care Measures in Children. *Ambulatory Pediatrics, 7*(1), 25-31

Erola, J., S. Jalonen and H, Lehti (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33-43.

Fallesen H. W. (2015). The educational performance of immigrant children: Examination of the native-immigrant education gap. *The Rockwool Foundation Research Unit.* Study Paper 97.

Fallesen, P. and F. Bernardi (2018). Parental welfare dependency and children's educational attainment in Denmark. *The Rockwool Foundation Research Unit.* Study Paper 126.

Gensowski, M., T. H. Nielsen, N. M. Nielsen, M. Rossin-Slater, and M. Wüst (2018). Childhood health shocks, comparative advantage, and long-term outcomes: evidence from the last Danish polio epidemic. *NBER Working Paper Series,* 24753.
Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153–161.

Gubta N. D. and S. Bengtsson (2017). Identifying the effects of education on the ability to cope with a disability among individuals with disabilities. *PLoS ONE, 12*(3), E0173659.

Gurney, J., K. Krull, N. Kadan-Lottick, H. Nicholson, P. Nathan, B. Zebrack, J. Tersak and K. Ness (2009). Social outcomes in the Childhood Cancer Survivor Study cohort. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 27*(14), 2390-5.

Gutstadt L. B., J. W. Gillette, D. A. Mrazek, J. T. Fukuhara, J. F. LaBrecque and R. C. Strunk (1989). Determinants of School Performance in Children With Chronic Asthma. *Am J Dis Child* 143(4):471–475.

Hald Andersen, Signe (2017). Social environmental insults, critical periods in brain formation and adverse outcomes in early adulthood. *The Rockwool Foundation Research Unit.* Study Paper no. 121.

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica (pre-1986), 47*(1), 153-161.

Hermann, C., N. Olivarius, A. Høst, K. Begtrup, and H. Hollnagel (2006). Prevalence, severity and determinants of asthma in Danish five-year-olds. *Acta Pædiatrica, 95*(10), 1182-1190.

Holmboe, O., J. Helgeland, K. Dahle, J. Clench-Aas and B. Van Roy, Betty (2006). Sosiale ulikheter I helse og bruk av helsetjenester blandt barn I Akerhus. *Rapport fra Kunnskapssenteret Nr 5*–2006

Joesch, J., and K. Smith (1997). Children's health and their mothers' risk of divorce or separation. *Social Biology, 44*(3-4), 159-169

Köhler-Forsberg, O. J., H. E. Sørensen, M. Nordentoft, J. McGrath, M. Benros, and L. Petersen (2018). Childhood Infections and Subsequent School Achievement Among 598,553 Danish Children. *The Pediatric Infectious Disease Journal, 37*(8), 731-737.

Lanari, M., S. Vandini, F. Adorni, F. Prinelli, S. Di Santo, M. Silvestri amd M. Musicco (2015). Prenatal tobacco smoke exposure increases hospitalizations for bronchiolitis in infants. *Respiratory Research, 16*(154), 152.

Lannerö, E., M. Wickman, G. Pershagen and L. Nordvall (2006). Maternal smoking during pregnancy increases the risk of recurrent wheezing during the first years of life (BAMSE). *Respiratory Research, 7*(1), 3.

Laursen, B., L. T. Nielsen, P. H. Christensen, H. Møller and B. Frimodt-Møller (2006) Børneulykker i Danmark. En registerbaseret analyse", Statens Institut for Folkesundhed, September 2006.

Laursen, B., H. Møller, and A. M. H. Toft (2008). Langvarige følgevirkninger hos voksne danskere af tilskadekomst ved ulykker, de var udsat for som barn, Jr. nr. 1.2006.26.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76* (3), 1071–1102.

Loft, L. (2011). Child Health and Parental Relationships. *International Journal of Sociology, 41*(1), 27-47.

Maslow, G., A. Haydon, C. Ford, and C. Halpern (2011). Young Adult Outcomes of Children Growing Up With Chronic Illness: An Analysis of the National Longitudinal Study of Adolescent Health. *Archives of Pediatrics & Adolescent Medicine, 165*(3), 256-261.

McInnes, K., C. L. Friesen, D. E. Mackenzie, D. A. Westwood, and S. G. Boe. (2017). Mild Traumatic Brain Injury (mTBI) and chronic cognitive impairment: A scoping review. *PLoS ONE, 12*(4), E0174847.

Ministry of Health (2016). https://www.sum.dk/English/~/media/Filer%20-%20Publikationer_i_pdf/2016/Healthcare-in-dk-16-dec/Healthcare-english-V16-dec.ashx

Moth, G., P. Schiotz, P. Vedsted (2008). A Danish population-based cohort study of newly diagnosed asthmatic children's care pathway – adherence to guidelines. *BMC Health Services Research, 8*(1), 130.

Nielsen, H. and S. Rangvid (2012). The impact of parents' years since migration on children's academic achievement. *IZA Journal of Migration, 1*(1), 1-23.

Prasad, M. R., P. Swank and L. Ewing-Cobbs (2017). Long-Term School Outcomes of Children and Adolescents With Traumatic Brain Injury. *Journal of Head Trauma Rehabilitation, 32*(1), E24-E32.

Reichman, N.E., H. Corman, and K. Noonan (2004). Effects of Child Health on Parents' Relationship Status. *Demography 41*(3), 569–84.

Reichman, N.E., H. Corman and K. Noonan (2008). Impact of Child Disability on the Family. *Maternal and Child Health Journal, 12*(6), 679-683.

Schmidt, M., S. Schmidt, J. Sandegaard, V. Ehrenstein, L. Pedersen and H. Sørensen (2015). The Danish National Patient Registry: A review of content, data quality, and research potential. *Clinical Epidemiology, 7*, 449-489.

Syse, A., T. Lyngstad and O. Kravdal (2012). Is mortality after childhood cancer dependent on social or economic resources of parents? A population-based study. *International Journal of Cancer, 130*(8), 1870-1878

Wilkinson, R. G. (1986). Socioeconomic differentials in mortality: interpreting the data on size and trends. *Class and Health: research and longitudinal data*. London: Tavistock, 1-20.

Appendix 1

Diagnosis groups and subgroups	Mean GPA	Std. dev. of GPA	# of obs.
Cancer			
Leukaemia/ALL	6.2	2.6	186
Leukaemia/AML	6.4	2.5	29
Leukaemia - other	6.1	2.2	53
Brain tumour	4.3	2.7	54
Lymphoma	6.4	2.4	40
Non-Hodgkin's lymphoma	7.2	4.5	4
Soft tissue Sarcoma	6.7	2.6	63
Renal cancer	6.4	2.3	39
Neuroblastoma	6.5	2.8	13
Hodgkin's lymphoma	6.4	2.4	37
Retinoblastoma	7.2	3.0	24
Liver cancer	4.8	2.9	13
Cancer - other	6.6	2.7	148
Muscular dystrophy and arthritis			
Arthritis	6.7	2.7	1,128
Arthrosis	6.4	2.7	155
Lupus	6.4	3.0	19
Osteoporosis	7.7	2.6	9
Calvé-Perthes' disease	6.0	2.9	552
Neurological disorders			
Meningitis	7.0	2.6	137
CNS diseases	5.8	2.7	274
Sclerosis	5.4	1.6	9
Epilepsy	5.8	2.7	4,260
TCI (small blood clot in the brain)	6.6	3.0	10
Cerebral palsy – brain injury	5.6	2.7	999
Apoplexy eller stroke	5.7	2.7	169
Benign brain tumour	5.8	2.4	104
Severe intracranial lesions	6.3	2.7	840
Diseases of inner organs			
Chron's disease	6.3	2.8	234
Ulcerative colitis	6.5	2.7	178
Chronic inflammation of the intestine	6.8	2.7	59
Chronic disease of the liver	6.0	2.9	44
Acute renal insufficiency	6.4	3.1	19
Chronic renal insufficiency	4.8	2.3	39
Renal insufficiency UNS	5.1	2.9	24
Hormonal and metabolic disorders			
Diabetes - type 1	6.5	2.7	1,297
Diabetes - type 2	5.3	3.1	23
Diabetes - other	5.9	2.8	66
Cystic fibrosis	6.3	2.4	88
Respiratory diseases			
Chronical bronchitis etc.	5.8	2.7	764
Chronic Obstructive Pulmonary Disease (COPD)	6.1	2.8	1,281
Traumatic injuries			
Severe fracture - spine, throat	6.2	2.6	628
Serious injury - spinal cord, nerves	5.9	2.6	251
Crushing Lesion	5.5	2.7	858
Traumatic Amputation	6.0	2.7	451
Severe combustion	5.7	2.7	315
Concussion	**6.3**	**2.7**	**18,350**

Appendix 2

Means of dependent variables – by children with and without illness

		Children with an illness	Children without illness	Differences
Child characteristics				
		--- Pct. ---	--- Pct. ---	
Gender	Female	44.2	49.1	-4.9***
	Male	55.8	50.9	4.9***
Ethnicity	Danish	93.8	92.8	1.0***
	Western countries	0.4	0.5	-0.1**
	Non-Western countries	5.8	6.7	-0.9
Siblings	First born	41.8	42.0	-0.3
	2nd born	39.4	38.5	0.9***
	3rd born	14.3	14.9	-0.6***
	4th born	3.5	3.5	0.0
	5th born or later	1.0	1.1	-0.1
Birth weight	Above 3000 grams	80.3	82.6	-2.2***
	2500-3000 grams	13.1	12.5	0.6***
	2000-2500 grams	3.8	3.3	0.6***
	1500-2000 grams	1.5	1.1	0.4***
	Below 1500 grams	1.3	0.6	0.6***
Parent characteristics				
Family status at birth	Living with both parents	91.0	93.3	-2.3***
	Living with one parent	8.8	6.6	2.2***
	Living with no parents	0.2	0.1	0.1***
Dad unknown		1.6	1.2	0.4***
		-- Mean --	-- Mean --	
Mothers age at birth		28.7	29.0	0.4***
		--- Pct. ---	--- Pct. ---	
Mothers Education at birth	High school or below	35.4	32.1	3.2***
	Vocational training	35.5	36.5	-1.0***
	Short higher education	3.5	3.5	0.0
	Medium higher education	18.2	19.9	-1.7***
	Long higher education	6.2	6.7	-0.5***
	Unknown	1.2	1.3	-0.1
	No data	0.0	0.0	0.0
Fathers Education at birth	High school or below	29.0	27.0	2.0***
	Vocational training	42.9	43.3	-0.4
	Short higher education	5.7	5.8	-0.1
	Medium higher education	10.2	11.3	-1.2***
	Long higher education	8.4	9.2	-0.8***
	Unknown	2.0	2.0	0.0
	No data	1.8	1.4	0.4***
Mothers occupation one year prior to birth	Employee, base level of qualification	37.5	37.9	-0.4
	Self employed	1.7	2.0	-0.2***
	Employee, highest level of qualification	6.8	7.6	-0.9***
	Employee, middle level of qualification	15.2	16.2	-1.1***
	Employee, other	4.4	4.1	0.3***
	Unemployed	13.7	13.0	0.7***
	Student	6.6	6.6	0.0
	Welfare claimant, early retirement	13.5	12.0	1.5***
	Unknown	0.6	0.7	-0.1

Fathers occupation one year prior to birth	Employee, base level of qualification	43.5	44.0	0.5*
	Self employed	6.2	6.3	-0.1
	Employee, highest level of qualification	11.3	12.4	-1.2***
	Employee, middle level of qualification	9.7	10.4	-0.7***
	Employee, other	8.8	8.1	-0.8***
	Unemployed	7.3	6.7	0.6***
	Student	3.7	3.6	0.0
	Welfare claimant, early retirement	7.4	6.7	0.7***
	Unknown	2.2	1.8	0.4***
Parent dead		3.2	2.2	1.0***
Observations		34,337	346,780	

Appendix 3

Dependent variable: GPA	Model 1.1	Model 1.2
Illness dummy	-0.257***	
Cancer		-0.216**
Muscular dystrophy and arthritis		-0.079
Neurological diseases		-0.585***
Inner organs		-0.237**
Hormonal and Metabolic		-0.103
Respiratory diseases		-0.212***
Traumatic injuries		-0.449***
Concussion		-0.177***
Birth year - 1994		
- 1995	-0.243***	-0.243***
- 1996	-0.070***	-0.069***
- 1997	-0.133***	-0.133***
- 1998	-0.146***	-0.146***
- 1999	0.130***	0.131***
Female	0.518***	0.518***
Ethnicity – Danish		
- Western countries	-0.085	-0.086
- Non-Western countries	-0.384***	-0.383***
First born		
- 2nd born	-0.503***	-0.503***
- 3rd born	-0.771***	-0.771***
- 4th born	-1.008***	-1.008***
- 5th born or later	-1.367***	-1.368***
Dad unknown	-0.494***	-0.493***
Parent dead	-0.388***	-0.388***
Mothers age at birth	0.333***	0.333***
Mothers age at birth - squared	-0.005***	-0.005***
Living with - both parents		
- One parent	-0.558***	-0.558***
- No parents	-0.807***	-0.802***
Mothers highest education – High school or below		
- Vocational training	0.299***	0.299***
- Short higher education	0.731***	0.731***
- Medium higher education	0.809***	0.809***
- Long higher education	1.266***	1.265***
- Unknown	-0.134***	-0.136***
- No data	0.565**	0.562**
Fathers highest education – High school or below		
- Vocational training	0.234***	0.234***
- Short higher education	0.614***	0.614***
- Medium higher education	0.826***	0.825***
- Long higher education	1.126***	1.125***
- Unknown	-0.149***	-0.150***
- No data	0.775***	0.778***
Mothers occupation - Employee, base level of qualification		
- Self employed	0.098***	0.097***
- Employee, highest level of qualification	0.422***	0.422***
- Employee, middle level of qualification	0.234***	0.234***
- Employee, other	-0.443***	-0.443***
- Unemployed	-0.318***	-0.319***

- Student	0.216***	0.215***
- Welfare claimant, early retirement	-0.326***	-0.326***
- Unknown	-0.219***	-0.219***
Fathers occupation - Employee, base level of qualification		
- Self employed	0.244***	0.244***
- Employee, highest level of qualification	0.631***	0.631***
- Employee, middle level of qualification	0.534***	0.534***
- Employee, other	-0.362***	-0.362***
- Unemployed	-0.273***	-0.273***
- Student	0.390***	0.391***
- Welfare claimant, early retirement	-0.233***	-0.233***
- Unknown	0.044	0.044
Birth weight – Above 3000 grams		
- 2500-3000 grams	-0.241***	-0.241***
- 2000-2500 grams	-0.271***	-0.270***
- 1500-2000 grams	-0.283***	-0.278***
- Below 1500 grams	-0.511***	-0.496***
Controls	Yes	Yes
Mean, dept. var.	6.513	6.513
R-squared	0.2150	0.2152
Observations	354,544	354,544

* Significant at 10%. ** Significant at 5%. *** Significant at 1%.

Notes: All regressions are estimated using OLS with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother’s age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, one or none), parental educational level and socioeconomic status prior to birth.

Appendix 4

Dependent variable: GPA	Model 2		
	All children	SE	Observations
Cancer – 0 days	0.197	[0.392]	45
- 1 day	-0.137	[0.493]	30
- 2-10 days	-0.191	[0.295]	83
- 10-100 days	-0.208	[0.137]	358
- Over 100 days	-0.417*	[0.224]	187
Muscular dystrophy and arthritis – 0 days	-0.0182	[0.073]	1.088
- 1 day	-0.0361	[0.153]	234
- 2-10 days	-0.243*	[0.135]	391
- 10-100 days	-0.174	[0.190]	150
- Over 100 days	0	[0.000]	0
Neurological diseases – 0 days	-0.510***	[0.052]	2.839
- 1 day	-0.476***	[0.088]	1.111
- 2-10 days	-0.647***	[0.071]	1.719
- 10-100 days	-0.976***	[0.114]	1.068
- Over 100 days	-1.386	[1.443]	65
Inner organs – 0 days	-0.408*	[0.244]	129
- 1 day	-0.0124	[0.251]	72
- 2-10 days	-0.0104	[0.174]	197
- 10-100 days	-0.443**	[0.175]	195
- Over 100 days	-0.279	[0.366]	4
Hormonal and Metabolic – 0 days	-0.279	[0.347]	49
- 1 day	-0.522	[0.382]	35
- 2-10 days	0.0456	[0.086]	787
- 10-100 days	-0.246**	[0.112]	592
- Over 100 days	-2.465***	[0.911]	11
Respiratory diseases – 0 days	-0.224**	[0.093]	752
- 1 day	-0.116	[0.106]	589
- 2-10 days	-0.268***	[0.104]	659
- 10-100 days	-0.422	[0.448]	44
- Over 100 days	-0.314***	[0.023]	0
Traumatic injuries– 0 days	-0.352***	[0.063]	1.536
- 1 day	-0.611***	[0.107]	565
- 2-10 days	-0.633***	[0.150]	278
- 10-100 days	-0.538**	[0.211]	124
- Over 100 days	0	[0.000]	0
Concussion – 0 days	-0.162***	[0.025]	9.629
- 1 day	-0.183***	[0.028]	7.835
- 2-10 days	-0.286***	[0.086]	875
- 10-100 days	0.0664	[0.941]	11
- Over 100 days	0	[0.000]	0
Controls	Yes		
Mean, dept. var.	6.513		
R-squared (adjusted)	0.216		
Observations	354,544		

* Significant at 10%. ** Significant at 5%. *** Significant at 1%.

Notes: The regression is estimated using OLS with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, one or none), parental educational level and socioeconomic status prior to birth.

Appendix 5

Dependent variable: GPA	Model 3		
	All children	SE	Observations
Cancer – age 0-2	-0.228	[0.196]	232
- Age 3-5	-0.487**	[0.209]	164
- Age 6-10	-0.0173	[0.209]	158
- Age 11-16	-0.120	[0.210]	149
Muscular dystrophy and arthritis – age 0-2	0.0276	[0.149]	260
- Age 3-5	-0.0360	[0.119]	453
- Age 6-10	-0.272**	[0.112]	513
- Age 11-16	-0.00289	[0.093]	637
Neurological diseases – age 0-2	-0.475***	[0.064]	2,261
- Age 3-5	-0.635***	[0.081]	1,362
- Age 6-10	-0.665***	[0.066]	1,947
- Age 11-16	-0.576***	[0.079]	1,232
Inner organs – age 0-2	-0.139	[0.475]	41
- Age 3-5	-0.297	[0.442]	32
- Age 6-10	-0.598***	[0.208]	142
- Age 11-16	-0.108	[0.124]	382
Hormonal and Metabolic – age 0-2	0.0769	[0.189]	187
- Age 3-5	-0.319*	[0.178]	208
- Age 6-10	0.0103	[0.108]	501
- Age 11-16	-0.177	[0.109]	578
Respiratory diseases – age 0-2	-0.219***	[0.063]	1,757
- Age 3-5	-0.0885	[0.173]	203
- Age 6-10	-0.145	[0.369]	56
- Age 11-16	-0.745*	[0.403]	29
Traumatic injuries– age 0-2	-0.330***	[0.109]	486
- Age 3-5	-0.143	[0.128]	366
- Age 6-10	-0.349***	[0.107]	581
- Age 11-16	-0.654***	[0.076]	1,070
Concussion – age 0-2	-0.145***	[0.033]	5,795
- Age 3-5	-0.145***	[0.037]	4,443
- Age 6-10	-0.216***	[0.038]	4,535
- Age 11-16	-0.218***	[0.042]	3,577
Controls	Yes		
Mean, dept. var.	6.513		
R-squared (adjusted)	0.216		
Observations	354,544		

* Significant at 10%. ** Significant at 5%. *** Significant at 1%.

Notes: The regression is estimated using OLS with heteroskedastic robust standard errors. We control for the following indicators: birth year, gender, ethnicity, birth order, birth weight, mother's age at birth, whether father is unknown, whether either of the parents is dead, family type at birth (living with both parents, one or none), parental educational level and socioeconomic status prior to birth.