I Z A

IZA DP No. 3646

The Role of Childhood Health for the Intergenerational Transmission of Human Capital: Evidence from Administrative Data

Martin Salm
Daniel Schunk

August 2008

The Role of Childhood Health for the Intergenerational Transmission of Human Capital: Evidence from Administrative Data

Martin Salm
*MEA, University of Mannheim
and IZA*

Daniel Schunk
IEW, University of Zurich

ABSTRACT

The Role of Childhood Health for the Intergenerational Transmission of Human Capital: Evidence from Administrative Data[*]

We use unique administrative German data to examine the role of childhood health for the intergenerational transmission of human capital. Specifically, we examine the extent to which a comprehensive list of health conditions – diagnosed by government physicians – can account for developmental gaps between the children of college educated parents and those of less educated parents. In total, health conditions explain 18% of the gap in cognitive ability and 65% of that in language ability, based on estimations with sibling fixed effects. Thus, policies aimed at reducing disparities in child achievement should also focus on improving the health of disadvantaged children.

Corresponding author:

Martin Salm
MEA
University of Mannheim
L13, 17
68131 Mannheim
Germany
E-mail: salm@mea.uni-mannheim.de

1. Introduction

Human capital in adulthood is to a large extent already determined during childhood. Measures of child development, such as cognitive and verbal ability, predict measures of human capital in adulthood, such as earnings and employment (Currie 2000, Currie and Thomas 1999, Dustmann et al. 2003), as well as commission of crime and partaking in other risky activities (Cunha and Heckman 2007, Heckman et al. 2006). However, child development is also strongly related to a child's socioeconomic background (Aughinbaugh and Gittleman 2003, Blau 1999, Carneiro et al. 2007, Taylor et al. 2004). This implies that many children from disadvantaged families fall behind early in life and find catching up later very difficult. Policies that aim to advance the skills and abilities of such children require detailed knowledge about pathways for the formation and intergenerational transmission of human capital (Currie 2008 and Heckman 2007).

In our study, we examine the role of childhood health as a pathway for this process. Ill health in childhood can adversely affect development in several ways (see surveys by Currie 2008 and Behrman 1996), which can be exacerbated by poor socioeconomic background: Pain, stress, and fatigue can reduce the ability to concentrate and to learn. Illness can crowd out other activities that might be beneficial to child development. Also, illness can change relationships between children, parents, and others in ways that might hurt cognitive and verbal development. Finally, some health conditions such as attention-deficit hyperactivity disorder or deafness, for example, can also have a direct, negative impact on cognitive or verbal ability, respectively. Previous studies such as Currie (2005, 2008) and Heckman (2007) emphasize that child health might be an important pathway for explaining disparities in child development between different socioeconomic groups. However, to the best of our knowledge, our study is the first to provide a comprehensive quantification of the extent to which gaps in child development between socioeconomic groups can be explained by differences in child health. Child health could indeed explain these developmental gaps if (1) child health conditions negatively impact child development, and (2) either the prevalence of negative child health conditions differs between socioeconomic groups or the magnitude of the effect of health conditions on child development differs between socioeconomic groups. In our study, we use unique administrative data for one German city to estimate the causal effect of a comprehensive list of child health conditions on young children's cognitive and verbal ability. We also use decomposition methods to estimate the extent to which gaps in cognitive and verbal ability by parental education groups can be

attributed to child health. We estimate both the share of the achievement gap attributable to differences in the prevalence of health conditions, as well as the share of the achievement gap attributable to differences in the magnitude of the effect of health conditions on child development.

Previous studies find a robust positive gradient between parents' socioeconomic status and child health for the United States (Case et al. 2002), but also for countries with universal health insurance coverage such as Canada (Currie and Stabile 2003) and the United Kingdom (Currie et al. 2007). The existing literature also provides some evidence for a negative correlation between a number of child health conditions and child outcomes. For example, Paxson and Schady (2007) find that health measures such as height for age and weight for age are positively related to language development in a study based on Ecuadorian data. Kaestner and Corman (1995) find that low birth weight, stunted growth, and severe diseases are correlated with lower reading and math abilities. However, a negative correlation between child health and child development does not prove a causal impact of child health on child development. Such a correlation could also be due to unobserved family characteristics and environmental factors, which could simultaneously influence both child health and child development. In order to account for such unobserved characteristics, Currie and Stabile (2006) use sibling fixed effects to examine the effect of child mental health conditions on child development. Ding et al. (2007) use genetic markers as instrumental variables in order to examine the effect of ADHD, depression, and obesity on older youths' school achievements.

Our study contributes to the existing literature on child health and child development in several ways. First, our data on child health are based on detailed examinations administered by government pediatricians during elementary school entrance medical exams. The information obtained in these exams is far more reliable than that based on survey questions administered to children's parents. Survey information is used in almost all previous studies, but it is well known that parents are not always fully informed about their children's health status and that the degree of knowledge about their children's health is strongly related to parent's socioeconomic status (see the survey by Currie 2000). Second, the fact that these exams are compulsory for all children in Germany at the age of six years gives the data an extraordinary degree of representativeness. Third, we examine a wider range of health conditions than those typically available in previous studies. We look at the effects of health conditions such as obesity, underweight, low birth weight, ear and eye conditions, mental health

4

conditions, asthma, and allergies; this allows us to capture the inherently multi-dimensional nature of health in our estimations. Fourth, we use sibling fixed effects models to address omitted variables bias. Sibling fixed effects models allow controlling for unobserved family characteristics that might influence both child health and cognitive and verbal development. Finally, and most importantly, our study goes beyond estimating the causal effect of child health conditions on child development by also quantifying the extent to which the achievement gap between parental education groups can be attributed to child health.

We find that that child health conditions have a substantially negative impact on child development. Moreover, child health conditions are more common among children of less educated parents, and their effect on developmental outcomes is more severe among children of less educated parents. In total, estimations with sibling fixed effects find that health factors account for 18.4% of the gap in cognitive ability between children of college educated parents and less educated parents, and for 64.8% of that in language ability. 10.8% of the cognitive ability gap and 12% of the language ability gap can be attributed to differences in the prevalence of health conditions, while 7.6% of the cognitive ability gap and 52.8% of the language ability gap are due to differences in the severity of the impact. Thus, both, differences in the prevalence as well as differences in the severity of the impact of health conditions contribute to the observed development gap. Our findings provide evidence that even in a country with almost universal health insurance coverage and a very generous system of child health prevention, health is still an important channel for the intergenerational transmission of human capital. Consequently, there could be much value in strengthening childhood programs that guarantee ongoing medical support through family based services or regular supervision, i.e. programs which strengthen home inputs that are complementary to professional medical treatment.

Our study continues as follows. Section 2 describes the data. Section 3 discusses the empirical strategy. Section 4 presents our estimation results, and section 5 concludes.

2. Data

Estimation sample. Our analysis is based on administrative data, which we obtained from the department of health services of the city of Osnabrueck, Germany. Osnabrueck is the third largest city in the German state of Lower Saxony, with a population of about 170,000. The data were collected during official school entrance medical examinations, which are compulsory for all

children at the age of 6 years, in the months before they enroll in elementary school. At the time of the examination, children in our sample had not previously attended elementary school. The examinations take between 50 and 70 minutes. The children are randomly assigned to one of three government pediatricians, who administer the examination. School entrance medical examinations in Osnabrueck have been increasingly standardized since the year 2000 (for details, see Rohling 2002), and the scope of the recorded data increased considerably beginning in 2002. From 2002 on, the data involve various medical and developmental tests, and information from each child's medical history which is contained in the child's mandatory "vaccination record" (Impfpass) and in the child's mandatory "health record" (Kinder-Untersuchungsheft). They also include information from two questionnaires on sociodemographic information and on the children's preschool experiences. One questionnaire was sent to parents before the examination, and parents answered the other questionnaire in a separate room while the child was being examined. Information based on these questionnaires was not available to the pediatrician during the exam.

Our data includes all regular 4,977 school entrance medical examinations, which took place in Osnabrueck in the years 2002 to 2005.[1] For 14.7% (N = 732) of the sample, self-reported information on parents' educational background was missing. Since parental education is used to stratify our sample by socioeconomic status, we exclude all 732 observations with incomplete information on parental education, meaning that the remaining estimation sample consists of 4,245 observations.[2] For 4.9% (N = 206) of the remaining sample, information on birth weight was missing, for 3% (N = 129) information on the mother's age was missing, for 2.1% (N = 87) information on preschool attendance was missing, and for 0.9% (N = 40) information on parents' employment (full-time employed, part-time employed, or unemployed) was missing. For the imputation of the missing information on birth weight, preschool attendance and parents' employment in the remaining sample, we employed a Markov chain Monte Carlo multiple imputation algorithm (Little and Rubin 2002, pp. 200-223), in which we condition on a maximum set of relevant information. The algorithm converges in less than 1,000 iteration steps, such that running the algorithm with different

---

[1] The department of health services stopped recording detailed sociodemographic information on the parents in 2006. Our data thus end in 2005.
[2] Information on the dependent variable *verbal ability* is missing for 11 additional children in the estimation sample. We dropped the observations on *verbal ability* from the sample in the corresponding analysis.

random seeds has virtually no effect on all estimation results reported in this paper.[3]

In addition to this full estimation sample, we also employ a subsample of 947 children where at least one sibling is also included in the estimation sample. This sample consists of 866 children with one sibling, 69 children with two siblings, and 12 children with three siblings. Information on siblings is available only if the sibling is also part of the estimation sample. This excludes siblings who started elementary school before 2002 or after 2005, as well as siblings who started elementary school in a different city, because those siblings did not take their school entrance exam in Osnabrueck. For both the full sample and the sibling sample, we defined two subsamples based on parents' education. For the full sample, 1,439 children had at least one parent with a college degree. For the remaining 2,806 children, neither of the parents had a college degree. For the sibling sample, 321 children had at least one parent with a college degree, and 626 children had no parents with a college degree.

Outcome variables. Outcome variables are measures of children's cognitive and verbal ability, since both pure cognitive skills as well as verbal skills are important predictors of children's school readiness and of their achievements later in life (Heckman 2006). Cognitive ability is assessed based on a CPM score. Raven's Coloured Progressive Matrices (CPM) is a test developed specifically for children and that measures abstract nonverbal reasoning ability (Raven et al. 1998). The test has been used successfully in a variety of cultural groups and cultural psychologists have studied it intensively (Reynolds and Kamphaus 2003). It consists of a sequence of colored patterns, and the tested child is required to fill in a missing pattern from a number of choices. The original colored matrices test consisted of 3 scales involving 12 items each. For the school entrance exam in Osnabrueck, 14 items were selected from the test, 4 from the first scale, 4 from the second scale, and 6 from the third scale. The overall CPM score is the sum of the results of the 14 items, i.e. 0 is the lowest possible score, and 14 is the highest possible score.[4]

---

[3] The imputation algorithm, the imputed versions of the dataset, and the corresponding estimation results are obtainable from the authors upon request. Note further that in addition to the procedure described above, which excludes observations with missing information on parents' educational background from the estimation sample, we also used an alternative procedure, where we also imputed the variables indicating parental education based on a Markov chain Monte Carlo algorithm. This allows us to use the full sample of 4,977 observations for the estimation. The results are similar to those reported in this paper and all conclusions of the paper are fully supported. Results are also similar for an analysis based on a sample with 3,811 observations. This sample involves no imputed data points, i.e. it is obtained after deleting all observations which had a missing value on at least one of the variables.

[4] As an alternative outcome variable we constructed a *standardized* CPM score with a mean of 100 and a standard deviation of 10, assuming a normal distribution of cognitive ability. We prefer

The mean score for the full sample is 10.86 with a standard deviation of 2.17 (see Table 1). The mean score for children of college educated parents is more than a third of a standard deviation higher than the mean score for children of less educated parents (11.37 versus 10.60), and this difference is highly significant. At 10.85, the mean score for the sibling sample is almost the same as for the full sample. However, differences between children of college educated parents and less educated parents are slightly, but insignificantly, larger for the sibling sample (11.56 versus 10.49). The correlation in the CPM scores between siblings is 0.31. Figure 1 illustrates the cumulative frequency distribution of CPM scores for the full sample. Across the entire distribution, the CPM score tends to be higher for children of college educated parents compared to children of less educated parents. Furthermore, we observe that almost all children answered at least some questions correctly. 10.1% of the children with college educated parents and 5.8% of the children with less educated parents answered all 14 questions correctly.

Our measure of verbal ability is a binary variable based on the pediatricians' assessment, whether a child's verbal ability corresponds to age level. The pediatricians' assessment of verbal ability is derived from a number of tests. Children are asked to describe what they see on a series of pictures, repeat actual and imaginary words and sentences, and talk freely about their favorite games and activities. The pediatrician assesses the accuracy of pronunciation and grammar as well as overall verbal ability. Verbal ability is judged to be at age level for 82.0% of children in the full sample. For children with college educated parents this share is 88.4%, as compared to 78.7% for children of less educated parents. The difference in these shares is statistically highly significant. In the sibling sample, verbal ability is slightly but insignificantly lower for all three groups (see Table 1). The correlation of verbal ability between siblings is 0.22.

Health variables. During the school enrollment examination, the pediatricians focus specifically on chronic health conditions, since only chronic conditions are relevant for the enrollment decision. Children with an acute illness that could influence their test results postponed the examination by a few weeks. In our analysis, we use detailed information on all chronic health conditions, which are available in our data. Information on health conditions comes from three sources: medical examinations during the school entrance examinations,

---

the specification with raw scores, which weights answers to all questions equally and does not artificially narrow differences between children with very low scores. However, estimation results (obtainable from the authors upon request) are similar and all conclusions from this paper are fully supported if the standardized outcome variable is used instead of the original CPM score variable.

children's official vaccination and health records, which parents are asked to bring to the exam, and information provided by the parents. During the examination, pediatricians measure the children's height and weight and examine their eyesight and hearing ability using standardized testing devices such as audiometers. Children above the 85th percentile of the body-mass-index distribution in our sample are classified as overweight, while those below the 5th percentile of the body-mass-index distribution in the full sample are classified as underweight. These definitions follow standard classifications from the U.S. Department of Health and Human Services (Center for Disease Control and Prevention, 2007). A binary variable on ear conditions takes on the value one if the examination with an audiometer reveals that the child's hearing abilities are below the normal range. A binary variable on eye conditions takes on the value one if there are limitations in visual ability. Visual exams are taken with glasses and include a test for the ability to see in three dimensions. Children who cannot distinguish red and green colors but who otherwise perform normally on all other visual tests are not classified as visually impaired. Information on birth weight comes from mandatory children's health certificate (Kinder-Untersuchungsheft), filled out by the attending pediatrician at the child's birth.[5] We create a binary variable for children with low birth weight (less than 2500 grams). This definition follows the International Statistical Classification of Diseases and Related Health Problems (World Health Organization 2007).

Information on mental health conditions is elicited from three sources. Before the child's examination, parents are asked to fill out a form, which includes questions on whether the child is hyperactive or has difficulties concentrating. During the child's examination, parents are asked to go to a separate room and to answer the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a standardized 25-item questionnaire designed for children from age four to eleven, and it asks questions about the child's emotional problems, pro-social and anti-social behavior, hyperactivity, and other behavioral problems (Goodman et al. 1998, Goodman 2001). Parents are reassured that their answers will not influence the physicians' school enrollment recommendation. The third and most important source of information comes from the pediatrician's observations during the 50 to 70 minutes of interaction with the child during the examination. Based on those three pieces of information, the pediatrician gives an overall assessment on whether a child has faces a mental health condition. The

---

[5] If the parents forgot to bring the child's health certificate to the examination, a missing value was recorded for birth weight. Self-reported information by the parents was not accepted.

pediatrician's assessment is likely to be more objective than parents' answers, since parents might have very different standards in judging their child's behaviors. We create a binary variable for mental health conditions according to the pediatrician's assessment.

Information on asthma and allergies is based on answers provided by parents in a questionnaire, on conversations with parents, and on the observation of children, as well as on stethoscopic examinations during the exam. We create a binary variable for children who suffer from asthma. Information on allergies is collected in the same way as for asthma. The definition of allergies includes allergic rhinitis and eczema.

Health conditions are quite common among young children in our sample (see Table 1). 14.8% of the children in the full sample are overweight and 4.7% are underweight. 5.7% were born with low birth weight. 11.0% are affected by ear problems and 19.3% by eye problems. 18.2% have face a mental condition. 6.4% of the children have asthma, and 3.5% suffer from allergies. Many health conditions are more common among children of less educated parents than among children of college educated parents. This relationship holds and is statistically significant for being overweight (17.7% vs. 9.1%), low birth weight (6.3% vs. 4.5%), mental conditions (20.7% vs. 13.2%), and for asthma (6.8% vs. 5.4%). For eye conditions (19.6% vs. 18.6%) and allergies (3.6% vs. 3.3%), prevalence is also higher for children of less educated parents, but the difference is not statistically significant. For underweight and ear conditions, prevalence is slightly but insignificantly lower for children with less educated parents. In the sibling sample, the prevalence of most health conditions is similar to the full sample. However, children in the sibling sample are less likely to be overweight (11.9% versus 14.8%) and much more likely to be low birth weight children (10.0% versus 5.7%). The high share of low birth weight children in the sibling sample can largely be explained by the twins in this sample, who tend to have considerably lower birth weight.

The incidence of health conditions tends to be positively correlated between siblings. Health conditions with the strongest positive correlations between siblings are low birth weight (0.48), overweight (0.26), asthma (0.15), and mental health conditions (0.16). These are also conditions with some of the largest differences in prevalence by parental education. The incidence of different health conditions also tends to be positively correlated for the same child, but correlation coefficients are quite low. We find the highest positive correlation coefficients for asthma and allergies (0.12), and for underweight and low birth weight (0.12). Other health conditions, which are significantly positively

correlated at the five percent level, are mental health conditions with underweight (0.05), low birth weight (0.04), ear conditions (0.06), eye conditions (0.05), and asthma (0.04), as well as ear conditions with eye conditions (0.05), and overweight with asthma (0.04).

Sociodemographic variables. We also include information on socioeconomic and demographic characteristics of children and their parents. We control for children's age, gender, preschool attendance (3+ years vs. less than 3 years), and type of family (child lives with both parents vs. other types). We also control for birth order by including the corresponding three dummy variables (fourth or later child is omitted category), and the number of siblings (one sibling is omitted category). These variables refer to the total number of siblings, not just the number of siblings included in the sample. We further control for the age of the mother at the time of the examination, for children who have at least one parent in full time employment, and for children with ethnic origins in Turkey, in Eastern European countries, or other foreign countries. Sample means and standard deviations for all included variables are listed in Table 1.

3. Empirical Strategy

Our aim is to estimate the causal effects of child health conditions on child development. Outcome variables (y_i) are the CPM score and the binary variable for verbal ability of child i. Explanatory variables include a vector of child health characteristics (H_i) and a vector of socioeconomic and demographic characteristics (X_i). We estimate both OLS models and sibling fixed effects models, in all cases with robust standard errors that are clustered at the family level. The estimation equation for the OLS model is:

$$y_i = \alpha_{edu} + H_i^{'}\beta_{edu} + X_i^{'}\gamma_{edu} + \varepsilon_i \tag{1}$$

β_{edu} and γ_{edu} are vectors of estimation coefficients. α_{edu} is the intercept and ε_i is a child-specific error term. Subscript *edu* refers to one of three estimation samples for which we estimate our models separately. The estimation samples are defined by parental education groups and consist either of children with college educated parents, or of children with less than college educated parents, or of the pooled sample of all children. β_{edu} measures the association between health conditions and development outcomes. This coefficient could differ between parental education groups, if for example children in one group receive better medical treatment, comply more with doctor's guidelines, or are in some other way better able to compensate the negative consequences of health conditions. In our analysis, we further control for a detailed list of socioeconomic and demographic

characteristics of children and their families (see Table 1). However, there could still be unobserved family-specific characteristics, such as for example parenting style, which directly impact both health conditions and child development.

Therefore, we also estimate sibling fixed effects models. These models control for unobserved family specific characteristics by examining the effect of differences between siblings' health on differences in siblings' cognitive or verbal development. The estimation equation for the fixed effects model is:

$$y_i = \alpha_f + H_i'\beta_{edu} + Z_i'\gamma_{edu} + u_i \qquad (2)$$

Z_i represents a vector of socioeconomic and demographic characteristics such as age, gender, preschool attendance, and birth order. The variables contained in Z_i are a subset of the variables in X_i, since Z_i excludes characteristics that are constant between siblings, such as parents' ethnic origin. Vector α_f represents the effect of unobserved family-specific influences for each family f. β_{edu} measures the causal effect of child health conditions on child development.

However, there are three sources of potential bias in the fixed effects specification which we discuss below. First, a bias could arise if siblings face systematic differences in available resources or parental treatment that affect both their health and development. Siblings in our data are close in age, implying the family's socioeconomic situation they experienced during childhood was similar. In our analysis, we also control for gender and birth order, two factors which could explain differential treatment. Furthermore, there is evidence that parents tend to treat children equally, for example with respect to dividing their wealth equally among children (Behrman and Rosenzweig 2004a). Existing evidence for Germany, based on the observation of transfer payments to children, shows that differential treatment of children is rare in Germany, and if it occurs, parents tend to spend more resources on disadvantaged children (e.g., Kuenemund et al 2005). Such compensating behavior could bias our estimates downward.

In this case, the strict exogeneity assumption could be violated, which is a necessary condition for estimating the causal effects β_{edu} consistently: $E(u_i \mid H_i, Z_i, \alpha_f)=0$. This assumption requires that the unobserved child specific component are not related to health conditions H_i or to socioeconomic characteristics Z_i. In our study we check the validity of the strict exogeneity assumption with a statistical test suggested by Wooldridge (2002, p. 285) based on the following equation:

$$\Delta y_{if} = \Delta H_{if}\beta_{edu} + \Delta Z_{if}\gamma_{edu} + \Delta u_{if} \qquad (3)$$

In this equation with first differences between two siblings $j \in \{1,2\}$ in family f, neither H_{1f} nor H_{2f} should be significant as additional explanatory variables. Consequently, to test the strict exogeneity assumption, we add H_{1f} (or H_{2f}, respectively) to the equation and carry out an F-test for the significance of H_{1f} (or H_{2f}, respectively).

This procedure is also a test for reverse causality, a second source of potential bias. However, given that our estimations focus only on chronic health conditions and only include young children who have never been to school, it is very unlikely that low cognitive or verbal achievement causes chronic health conditions. At the age of the children in our study, needs for medical or psychological treatments are diagnosed and initiated by the parents rather than the children, and parents play the key role in the supervision of those treatments. Similarly, it is unlikely that poor child health reduces parental education, considering that the average age of mothers in our sample is above 36 years. In fact, this argument is supported by estimations including only children of mothers who were older than 25 years at child birth.

The third potential source of bias arises from measurement errors in explanatory variables. Sibling fixed effects estimation can exacerbate the measurement error problem by increasing noise due to potentially strong correlations between siblings and reducing exogenous variation. Specifically, Griliches (1979) and Bound and Solon (1999) show that measurement errors can result in substantial downward bias of coefficients in models with sibling fixed effects. However, in our study biases arising from this source seem much less of a problem than in existing studies for the following two reasons. First, there is substantial variation in health conditions between siblings, and second, our measures for health conditions are derived from administrative data, which were collected by experienced physicians in highly standardized procedures. There is much less scope for measurement error in our health variables than in previous studies on child health, which use data based on survey questions, and particularly less than in the studies discussed by Griliches (1979) and Bound and Solon (1999), which estimate the returns to education on wages.

Having estimated both OLS and fixed effects specifications separately for all three samples, we proceed by examining how much of the child achievement gap can be attributed to differences in the prevalence and severity of the impact of negative health conditions. For this purpose we use decomposition analysis, similar to Oaxaca (1973). We first calculate the part of the difference which can be attributed to differences in the prevalence of health conditions. For this calculation, we keep the magnitude of the effect of health conditions on child

development as well as all non-health factors constant between education groups. The prevalence effect is then calculated by the formula below:

$$prevalence\text{-}effect = (\overline{H}_{college} - \overline{H}_{nocollege})' \beta_{pooled} \qquad (4)$$

Here, $\overline{H}_{college}$ is the vector of mean values for health conditions for the sample with college educated parents, while $\overline{H}_{nocollege}$ is the corresponding vector for the sample with less educated parents. β_{pooled} is the vector of estimation coefficients for the reference group. Following Neumark (1988) we take estimation coefficients for the pooled sample as the reference group.

We also calculate the extent to which the achievement gap by parental education can be attributed to differences in the magnitude of the effect of health conditions on child development, i.e. we calculate the severity effect. The magnitude of the effect of health conditions for the sample with college educated parents is measured by $\beta_{college}$, while the magnitude of this effect for the sample with less educated parents is measured by $\beta_{nocollege}$. In order to calculate the impact of differences in the magnitude of health effects, we keep the prevalence of health conditions as well as all non-health factors constant between parental education groups. The severity effect is then calculated by the following formula:

$$severity\text{-}effect = \overline{H}'_{pooled}(\beta_{college} - \beta_{nocollege}) \ . \qquad (5)$$

The vector $\overline{H}_{pooled}$ refers to the prevalence of health conditions for the pooled sample. The sum of the prevalence effect and the severity effect constitutes the total effect of child health conditions on the achievement gap.

4. Results
4.1. Regression results for CPM score

Table 2 shows estimation results for CPM scores, which measure cognitive ability. Panel A of Table 2 presents results for an OLS model for the full estimation sample, including children with and without siblings. The results for the pooled sample, which includes both parental education groups, indicate a strong negative relationship between most health conditions and cognitive ability (see column 1). CPM scores for overweight children are on average 0.32 points lower than for the comparison group, i.e. children with normal body mass index. In contrast, underweight children do not differ significantly in cognitive ability from children with normal BMI. Low birth weight is strongly linked with lower cognitive ability. Children with a birth weight of less than 2500 grams score on average 0.78 points less on the CPM scale, this is more than a third of a standard

deviation. Eye conditions have a negative impact on cognitive ability. Eye conditions can reduce the ability to recognize pictures, symbols, and patterns. Among the health conditions examined in this study, mental health conditions have the strongest negative impact on cognitive ability. CPM scores for children with a mental health condition are reduced by 1.07 points, which is equivalent to about half of a standard deviation. Mental health conditions such as attention deficit hyperactivity disorder reduce children's ability to concentrate and focus on one task. Cognitive ability is also significantly reduced for children with asthma. Asthma is the leading cause of children's trips to the emergency room and hospitalization (Currie 2005). If not treated adequately, asthma can severely limit children's activity. However, if controlled well, asthma attacks can be prevented, and asthma should have little or no direct effect on cognitive development (Currie 2005). Finally, we find that children with allergies do not differ significantly in their cognitive ability from those without allergies. Overall, our estimation results for the relationship between health conditions and cognitive ability are in line with results from existing studies.[6]

Column 2 of Panel A shows estimation results for a subsample of children with at least one college educated parents. Compared to the pooled sample, estimation coefficients tend to be less negative, and coefficients are significant for three health conditions only: low birth weight, eye conditions, and mental health conditions. The coefficient for low birth weight is even more negative than in column 1. One explanation for this finding is a higher prevalence of children with very low birth weight in the sample with college educated parents, possibly because those children are more likely to survive. The coefficient for eye conditions is of similar size compared to the full sample, implying that the relationship between poor eyesight and cognitive ability does not vary by parental education group. Likewise, the coefficient for mental health conditions is of similar size compared to the pooled sample. This result is in line with Currie and Stabile (2006), who also find that the negative impact of mental health conditions on cognitive ability is not ameliorated by higher parental education; this could reflect the still limited effectiveness of treatment options for child mental health conditions. Compared to the pooled sample, there is no significantly negative effect for overweight and asthma in the subsample of

---

[6] Previous studies also find a negative relationship between childhood health conditions and child cognitive abilities for obesity (Ding et al. 2007), low birth weight (for example Behrman and Rosenzweig 2004b), asthma (Halterman et al. 2001), and also a particularly strong relationship between child development and mental health conditions (Currie and Stabile 2006). Paxson and Schady (2007) also find no significant relationship between low BMI and cognitive ability of Ecuadorian children. A number of studies from different scientific fields discuss the effects of specific health conditions on child development. Currie (2008) provides an excellent overview of the current state of the literature.

children with college educated parents. These findings indicate that children of college educated parents can somehow compensate for the negative consequences of overweight and asthma.

Column 3 of Panel A shows results for the sample with less educated parents. The relationship between health conditions and cognitive ability scores is stronger compared to the pooled sample and to the sample with college educated parents, and estimation coefficients are more negative. We find a significantly negative association with cognitive ability for the following health conditions: overweight, low birth weight, ear conditions, mental health conditions, and asthma. Among these conditions, coefficients are more negative than for the sample with college educated parents for all health conditions except for eye conditions and allergies. In order to test whether differences in estimation coefficients between the two samples of children with college educated parents and children with less educated parents are statistically significant, we further estimate a model where all health conditions are fully interacted with parental education level and we test for the statistical significance of the interaction terms. The difference in coefficients is statistically significant only for asthma (p = 0.006).[7]

Panel B of Table 2 shows OLS results for the sample of children with at least one sibling in our estimation sample. The estimation specification is identical to Panel A, but the sample size is reduced from 4,245 observations to 947 observations for the pooled sample. Due to the smaller sample size, coefficients are now measured less precisely and many estimates are no longer statistically significant, even though point estimates are similar. For the pooled sample shown in column 1, only the coefficients for low birth weight and for mental health conditions are statistically significant. For the sample with college educated parents (column 2), mental health conditions and low birth weight are associated with significantly lower cognitive ability scores. For the sample of children with less than college educated parents (column 3), there is a significantly negative relationship between cognitive ability and overweight, low birth weight, and mental health conditions. The difference between coefficients in column 2 and column 3 is significant only for overweight (p=0.081).

Panel C shows fixed effects estimation results which control for unobserved family characteristics by comparing cognitive ability scores for siblings with and without health conditions. The sample is identical to the sample in Panel B. Results for the pooled sample are shown in column 1. Among health conditions, we find negative coefficients for low birth weight, underweight,

---

[7] This analysis is not shown in the tables, but can be obtained from the authors upon request.

mental conditions, and for asthma, but only mental health conditions have a significantly negative impact on cognitive ability. The effects of mental health conditions on cognitive ability are highly significant for both the sample with college educated parents (column 2) and for the sample with less than college educated parents (column 3). Finally, in line with the pattern observed in the OLS estimation, we find that asthma has a large negative effect on cognitive ability for children of less educated parents, but the coefficient is insignificant and close to zero for the sample with college educated parents.

The fixed effects estimators rely on the strict exogeneity assumption that the child specific error term is not related to observed characteristics. In order to test for the validity of this assumption we employ the test statistic described in section 3, and find no evidence that the strict exogeneity assumption might be violated or that reverse causality might play a role (p-value = 0.686).

4.2. Regression results for verbal ability

Our estimation results for the effect of health conditions on verbal ability are shown in Table 3. We estimate linear probability models with a binary indicator as dependent variable which takes on the value one if a child's verbal ability is at age level, and zero otherwise. Again, panel A shows results based on the full sample, which includes both children with siblings and children without siblings in our sample. Column 1 shows estimation results for the pooled sample, i.e. both children with college educated parents and children with less educated parents. Among the health conditions examined in our study, only low birth weight and asthma have a significant effect on verbal ability. Low birth weight reduces the probability of at age level verbal ability by 5.6%. Children with asthma have an 8.6% lower probability that their verbal ability is at age level. All other health conditions have no significant effect on verbal ability. For children with college educated parents, none of the health conditions has a significant effect on verbal ability (column 2). The coefficient for asthma is much smaller than in the pooled sample. For children of less than college educated parents, only asthma has a significantly negative effect on verbal ability, and the magnitude of this effect is higher than in the pooled sample (column 3).

Panel B shows estimation results for the subsample of children with at least one sibling in our dataset. Otherwise, the specification is exactly as in Panel A. Verbal ability is reduced by low birth weight and mental health conditions (column 1). For children with high parental education, only allergies have a significantly negative impact on verbal ability (column 2). For children with less

educated parents, the coefficient for mental health conditions is of similar size compared to the pooled sample, but not statistically significant (column 3).

Panel C shows estimation results for the fixed effects specification, which examines the effect of health disparities between siblings on differences in verbal ability. There are significant effects for mental health conditions for the pooled sample in column 1. While no health condition is significantly related to language ability for the sample with college educated parents, mental health conditions have a significantly negative effect on cognitive ability for the sample with less than college educated parents (column 3).

As for the investigation of cognitive ability, we test for the validity of the strict exogeneity assumption of the child specific component of the error term with the test described in section 3. The strict exogeneity assumption is not violated (p-value = 0.540).

4.3. Decomposition analysis

The results discussed above suggest that health conditions can have a negative effect on child development. But to what extent can the observed achievement gap be attributed to child health? Decomposition analysis reveals what share of this gap can be explained by child health conditions. Table 4 shows the results for the decomposition analysis for cognitive ability (Panel A) and for verbal ability (Panel B). The third row of this table shows that the total difference in mean CPM scores between parental education groups is 0.771 points for the full sample and 1.057 points for the sibling sample, about half a standard deviation of the CPM score.

The prevalence effect defines the share of the cognitive ability gap that differences in the prevalence of health conditions can explain. It accounts for 0.129 points of the gap in the OLS analysis for the full sample, 0.201 points of the gap in the OLS analysis for the sibling sample and 0.114 points in the fixed effects analysis. As a share of the total difference, these numbers amount to 16.7%, 19.0%, and 10.8%, respectively. We calculate standard errors for the prevalence effect based on 200 bootstrap replications, and we find that the prevalence effect is significantly different from zero for all three specifications.

The severity effect explains which share of the cognitive ability gap can be accounted for by differences in the magnitude of the effect of health conditions on cognitive ability. It accounts for an additional 12.2% of the achievement gap for the OLS analysis based on the full sample, 15.8% for the OLS analysis based on the sibling sample, and 7.6% for the fixed effects analysis. These effects are slightly lower in magnitude than the prevalence effects. They are measured less

precisely and are not statistically significant. The total effect of child health conditions on cognitive ability is calculated as the sum of the prevalence and the severity effects. In total, child health conditions account for 28.9% of the cognitive ability gap for the OLS analysis based on the full sample, 34.8% for the OLS analysis based on the sibling sample, and 18.4% for the fixed effects analysis.

Panel B shows the decomposition analysis for verbal ability. The difference in the fraction of children whose verbal ability is at age level amounts to 9.7 percentage points between parental education groups for the full sample, and to 10.8 percentage points for the sibling sample. As share of the verbal ability gap, the prevalence effect can account for 5.2% for the OLS estimation based on the full sample, 11.1% for the OLS estimation based on the sibling sample, and 12% for the fixed effects estimation. Our results are only statistically significant for the OLS estimation based on the full sample. The severity effect amounts to 17.5% for the OLS analysis based on the full sample, 19.5% for the OLS analysis based on the sibling sample, and 52.8% for the fixed effects analysis. Because of large standard errors, these effects are not statistically significant. The total effect of child health conditions on verbal ability accounts for 22.7% of the verbal ability gap based on the OLS analysis for the full sample, 30.6% based on the OLS analysis based on the sibling sample, and 64.8% based on the fixed effects analysis.

Overall, our findings from the decomposition analysis suggest that health conditions are a very important factor in the intergenerational transmission of human capital. In a back-of-the-envelope calculation, Currie (2005) suggested that child health might explain around 12% of the achievement gap between white and black children in the United States at school entry. Our results suggest that the influence of health might be even stronger.

4.4 Juhn-Murphy-Pierce decomposition

We demonstrated above that child health conditions explain a substantial share of the achievement gap. Both differences in the prevalence of health conditions as well as differences in the severity of the effects of health conditions contribute to differences in achievement. One question we have not yet addressed is how these effects vary along the achievement distribution. Can health conditions explain more of the achievement gap at the lower end or at the higher end of the achievement distribution?

Our answer to this question employs a nonparametric decomposition method introduced by Juhn, Murphy, and Pierce (1993), which allows comparing

the size of the prevalence effect and the size of the severity effect across the entire ability distribution. According to this method, observed outcomes can be described as a combination of the following components

$$y_i = H_i \beta_{edu} + X_i \gamma_{edu} + \alpha_{edu} + \varepsilon_i \qquad (6)$$

Here, y_i is the observed cognitive ability score of child i, H_i is a vector with information on health conditions, X_i is a vector with socioeconomic and demographic information and ε_i is the component of cognitive ability accounted for by unobservable characteristics. α_{edu} is an intercept, and β_{edu} and γ_{edu} are vectors of coefficients which can vary between parental education groups $edu \in \{college, nocollege\}$. We can think of the last three terms of the equation above as involving two components: The distribution of the non-health related component of the cognitive ability equation, F_{edu}, and each individual's percentile in this distribution function, $\theta_{i,edu}$. By definition of the cumulative distribution function, we have:

$$X_i \gamma_{edu} + \alpha_{edu} + \varepsilon_i = F_{edu}^{-1}(\theta_{i,edu}) . \qquad (7)$$

Here, $F_{edu}^{-1}(.)$ is the inverse of the cumulative distribution function of non-health characteristics for children in parental education group edu. The decomposition method by Juhn et al. (1993) is based on the creation of two hypothetical distributions of cognitive ability. The first of these hypothetical distributions allows for variation in individual health characteristics, but keeps everything else constant between parental education groups. Thus, we construct hypothetical outcome measures where the size of the effect of health characteristics on cognitive ability and the distribution of the non-health component of cognitive ability is constant between parental education groups:

$$y_i^1 = H_i \beta_{pooled} + F_{pooled}^{-1}(\theta_{i,edu}) , \qquad (8)$$

where β_{pooled} is the vector of coefficients for health conditions based on the estimation with the pooled sample. $F_{pooled}^{-1}(.)$ is the inverse of the cumulative distribution function for the non-health component of cognitive ability. This component refers to the distribution in the pooled sample, while the term $F_{edu}^{-1}(.)$ in equation (7) refers to the distribution in the sample of each child's parental education group. For example, a child in the 75th percentile for this component in the group with college educated parents is now assigned the value for the 75th percentile for this component in the pooled sample. The second hypothetical distribution also keeps the distribution of the non-health related component of

cognitive ability constant between parental education groups, but allows for differences in the severity of the effect of health conditions on cognitive ability in addition to differences in the prevalence of health conditions:

$$y_i^2 = H_i \beta_{edu} + F_{pooled}^{-1}(\theta_{i,edu}) \qquad (9)$$

The observed difference in cognitive ability scores at quantile θ of the cognitive ability distribution, Δy_θ, can now be decomposed as follows:

$$\Delta y_\theta = \Delta y_\theta^1 + (\Delta y_\theta^2 - \Delta y_\theta^1) + (\Delta y_\theta - \Delta y_\theta^2) \qquad (10)$$

The first term on the right hand side of the equation above represents the part of the cognitive ability gap that can be explained by differences in the prevalence of health conditions. The second part represents the part that can be explained by differences in the severity of the effect of these health conditions. And the third part represents differences that can be explained by differences in socioeconomic and demographic as well as unobserved characteristics.

Figure 2 shows the distribution of the cognitive ability gap across the entire cognitive ability distribution.[8] The difference is highest for the lowest quantiles and decreases for higher quantiles. There is no difference at the top end of the distribution, since the most gifted children of both groups score the maximum of 14 points on the CPM scale. The distribution of the prevalence effect is shown in Figure 3. The prevalence effect is positive except for the very lowest quantiles of the cognitive ability distribution and for some quantiles around the 30th quantile, and it exhibits a roughly increasing pattern. The distribution of the severity effect is illustrated in Figure 4. Again, this effect is mostly positive except for the lowest quantiles and for some quantiles around the 30th quantile, and the severity effect is largest for the highest quantiles of the cognitive ability distribution. The residual effect depicted in Figure 5 shows a decreasing pattern, similar to the distribution of the total difference in cognitive ability scores.

In summary, both differences in the prevalence of health conditions and differences in the severity of the effect of health conditions contribute to the cognitive ability gap across most of the ability distribution, with the exception of the lowest part of the distribution. The figures show that the component of the cognitive ability gap which can be attributed to health conditions is largest at the higher end of the cognitive ability distribution. This indicates that health conditions are more likely to hold back the gifted children of less educated parents than their peers with college educated parents.

---

[8] The distributions, on which the graph is based, are smoothed, assuming that the ability of children scoring p points on the CPM-scale is equally distributed between p points and p+1 points.

5. Conclusions

We use unique administrative data from German elementary school medical entrance examinations in order to examine the role of child health in the intergenerational transmission of human capital. These examinations are mandatory for all children at the age of six years and therefore have an extraordinary degree of representativeness. Our data contain professionally measured information on child health conditions such as obesity, underweight, low birth weight, ear problems, eye problems, behavioral problems, asthma, and allergies. This allows us to capture the inherently multi-dimensional nature of health in our estimations. We find large differences in cognitive and verbal ability by parental education groups, and we also find that child health conditions are more common among children of less educated parents. Qualitatively, our main results are first that child health conditions have a substantially negative effect on child development. Second, differences in both the prevalence as well as the severity of child health conditions contribute to developmental gaps between socioeconomic groups.

Our findings also quantify the importance of health as a pathway for the intergenerational transmission of human capital: Based on estimations with sibling fixed effects, we find that 18.4% of the gap in cognitive ability and 64.8% of that in verbal ability between children of college educated parents and less educated parents can be attributed to health conditions. 10.8% of the cognitive ability gap and 12% of the language ability gap can be attributed to differences in the prevalence of health conditions, while 7.6% of the cognitive ability gap and 52.8% of the language ability gap can be attributed to differences in the severity of the impact of health conditions.

The gist of our findings can be nicely illustrated with the example of asthma. There is a growing body of evidence that common health conditions, such as asthma, need not have negative consequences on child development if they are well controlled (see, e.g., Halterman et al. 2001). This is the case for our subsample with college educated parents, for which we find no negative effect of asthma on children's cognitive or verbal abilities. However, we find robust evidence for a negative effect of asthma on cognitive ability for the subsample with less educated parents.

The extent to which health contributes to the developmental gap might seem surprising given that the German health system is characterized by almost universal health insurance coverage and a focus on child health and prevention programs: 99.8% of the German population are enrolled in mandatory health insurance, and those who are not enrolled are mostly the very rich (German

Federal Statistical Office 2004). Since 1971, there has been a child health prevention program which involves nine pediatric examinations between birth and the age of five years. All examinations as well as all subsequent treatment costs are paid for by the German health insurance system – without any deductible. While attendance of the first seven examinations of these examinations is well above 90%, attendance of the ninth examination at the age of five years is only about 80% (Schubert et al. 2004). Moreover, there is evidence that socioeconomic status strongly determines attendance. Children with parents from lower socioeconomic segments of the population attend these examinations significantly less than children from higher socioeconomic groups (Klocke 2001). In view of these institutional facts, the share of the ability gap that we have attributed to differences in health conditions can not only be explained by differences in access to medical care. They imply a need for further policy support to remedy the impact of family background on the ability to address these health conditions appropriately. More specifically, our results indicate that policies should strengthen childhood programs that guarantee ongoing medical support through family based services or regular supervision programs. There is a need for programs which particularly strengthen the home inputs complementary to professional medical treatment. Possible further interventions could include mandatory health examinations in preschools and health training for preschool supervisors and parents of small children thus enabling them to better recognize and address children's health problems.

In summary, our findings confirm that physical, mental and cognitive conditions are closely linked. We show that it is important for policy to better understand the lifecycle of skill and health formation. Our findings suggest that policies aimed at reducing disparities in child development should also aim at reducing disparities in health with a focus on children from disadvantaged families.

References

Aughinbaugh, A. and M. Gittleman (2003): Does Money Matter? A Comparison of the Effect of Income on Child Development in the United States and Great Britain. *The Journal of Human Resources*, 38, 2, 416-440.

Behrman, J. (1996): The impact of health and nutrition on education. *The World Bank Research Observer,* 11, 23-37.

Behrman, J. and M. Rosenzweig (2004a): Parental allocations to children: New evidence on bequest differences among siblings. *Review of Economics and Statistics*, 86, 637-640.

Behrman, J. and M. Rosenzweig (2004b: Returns to birthweight. *Review of Economics and Statistics*, 86, 586 – 601.

Blau, D. (1999): The effect of income on child development. *Review of Economic and Statistics*, 81, 261-276.

Bound, J. and G. Solon (1999): Double trouble: on the value of twin-based estimation of the return to schooling. *Economics of Education Review*, 18, 169-182.

Carneiro, P., C. Meghir, and M. Parey (2007): Maternal education, home environments and the development of children and adolescent, IZA Discussion Paper 3072.

Case, A., D. Lubotsky, and C. Paxson (2002): Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92, 1308-1334.

Case, A., A. Fertig, and C. Paxson (2005): The Lasting Impact of Childhood Health and Circumstance. *Journal of Health Economics*, 2005, 24, 365-389.

Center for Disease Control and Prevention (2007): U.S. Department of Health and Human Services. http://www.cdc.gov/nccdphp/dnpa/bmi/childrens˙BMI/about˙childrens˙BMI.htm#How%20is%20BMI%20calculated; accessed May 7, 2007.

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

Currie, J. (2000): Child health in developed countries. In: *Handbook of Health Economics,* Vol. 1B, edited by A. Culyer and J. Newhouse. North Holland.

Currie, J. (2005): Health disparities and gaps in school readiness, in: The Future of Children: School Readiness: Closing Racial and Ethnic Gaps. 15, 1, 117-138.

Currie, J. (2008): Healthy, Wealthy and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. Forthcoming: *Journal of Economic Literature*.

Currie, J. and M. Stabile (2003): Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review*, 93, 1813 – 1823.

Currie, A., M. Shields, and S. Price (2007): Is the child health / family income gradient universal? Evidence from England. *Journal of Health Economics*, 26, 212-232.

Currie, J. and M. Stabile (2006): Child Mental Health and Human Capital Accumulation: the Case of ADHD. *Journal of Health Economics*, 25, 6, 1094-1118.

Currie, J. and D. Thomas, (2001): Early Test Scores, Socioeconomic Status, School Quality and Future Outcomes. *Research in Labor Economics*. 20, 103-132.

Ding, W., S. Lehrer, J. Rosenquist, and J. Audrain-McGovern (2007): The impact of poor health on education: new evidence using genetic markers. Working Paper, Queens University.

Dustmann, C., N. Rajah, and A. van Soest (2003): Class size, education and wages. *Economic Journal*, 113, 99-120.

Ermisch, J and M. Francesconi (2000): Educational Choice, Families and Young People's Earnings. *Journal of Human Resources*, 35, 1, 146-176.

German Federal Statistical Office (2004): Krankenversicherung in Deutschland. Sonderauswertung des Mikrozensus 2003. Wiesbaden 2004.

Goodman, R. (2001): Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.

Goodman, R., H. Meltzer, and V. Bailey (1998): The Strengths and Difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, 7, 125-130.

Griliches, Z. (1979): Sibling models and data in economics: beginnings of a survey. *Journal of Political Economy*, 87, S37-S64.

Halterman, J., G. Montes, C. Aligne, J. Kaczorowski, A. Hightower, and P. Szilagyi (2001): School Readiness among Urban Children with Asthma. *Ambulatory Pediatrics,* 1, 4 , 201–05.

Heckman, J. (2006): Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 312, 1900 – 1902.

Heckman, J. (2007): The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences.* 104, 13250–13255.

Heckman, J., J. Stixrud, and S. Urzua (2006): The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24, 3, 411-482.

Juhn, J., K. Murphy, and B. Pierce (1993): Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101, 410-442.

Kaestner, R. and H. Corman (1995): The impact of child health and family inputs on child cognitive development. NBER Working paper 5357.

Klocke, A. (2001): Armut bei Kindern und Jugendlichen und die Auswirkungen auf die Gesundheit. Gesundheitsberichterstattung des Bundes (GBE) Heft 03/01. Verlag Robert-Koch-Institut, Berlin.

Kuenemund, H., A. Motel-Klingebiel, and M. Kohli (2005): Do Intergenerational Transfers From Elderly Parents Increase Social Inequality Among Their Middle-Aged Children? Evidence from the German Aging Survey. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* 60, S30-S36.

Little, R.J.A. and D.B. Rubin (2002): Statistical Analysis with Missing Data (2nd edition). Wiley, New York.

Neumark, D. (1988): Employers' discriminatory behaviour and the estimation of wage discrimination. *Journal of Human Resources*, 23, 279-295.

Oaxaca, R. (1973): Male-female wage differentials in urban labor markets. *International Economic Review*, 14, 693 – 709.

Paxson, C. and N. Schady (2007): Cognitive development among young children in Ecuador. *Journal of Human Resources*, 42, 49-84.

Raven, J., J.C. Raven, and J.H. Court (1998): Couloured Progressive Matrices. Oxford Psychologists Press.

Reynolds, C.R. and R.W. Kamphaus, Eds. (2003): Handbook of Psychological and Educational Assessment of Children: Intelligence, Aptitude, and Achievement (2nd edition). Guilford Press.

Rohling, I. (2002): Gesundheit und Entwicklungsstand der Osnabruecker Schulanfaenger. Stadt Osnabrueck, Fachbereich Soziales und Gesundheit.

Schubert, I. and K. Horch (2001): Schwerpunktbericht der Gesundheits-berichterstattung des Bundes. Gesundheit von Kindern und Jugendlichen. Verlag Robert-Koch-Institut, Berlin.

Taylor, B.A., E. Dearing, and K. McCartney (2004): Incomes and Outcomes in Early Childhood. *Journal of Human Resources*, 39, 4, 980-1007.

Wooldridge, J. (2002): Econometric Analysis of cross section and panel data. MIT Press.

World Health Organization (2007): International Statistical Classification of Diseases and Related Health Problems – Version for 2007. http://www.who.int/classifications/apps/icd/icd10online/?gp05.htm+p05, accessed May 7, 2007.

Table 1: Sample Means

Variable	Full sample				Sibling sample			
	Pooled sample	Parents college	Parents no college	≠ sign.	Pooled sample	Parents college	Parents no college	≠ sign.
CPM score (0-14)	10.857	11.368	10.596	***	10.850	11.555	10.489	***
	(2.172)	(2.012)	(2.204)		(2.203)	(1.877)	(2.271)	
Verbal ability at	0.820	0.884	0.787	***	0.800	0.872	0.763	***
age level	(0.385)	(0.321)	(0.410)		(0.400)	(0.335)	(0.426)	
Overweight	0.148	0.091	0.177	***	0.119	0.062	0.149	***
	(0.355)	(0.288)	(0.382)		(0.324)	(0.242)	(0.356)	
Underweight	0.047	0.048	0.046		0.052	0.053	0.051	
	(0.211)	(0.214)	(0.210)		(0.222)	(0.224)	(0.220)	
Low birth weight	0.057	0.045	0.063	**	0.100	0.081	0.110	
	(0.231)	(0.208)	(0.243)		(0.301)	(0.273)	(0.313)	
Ear condition	0.110	0.117	0.106		0.116	0.143	0.102	*
	(0.313)	(0.321)	(0.308)		(0.321)	(0.351)	(0.303)	
Eye condition	0.193	0.186	0.196		0.196	0.162	0.214	*
	(0.395)	(0.389)	(0.397)		(0.398)	(0.369)	(0.411)	
Mental condition	0.182	0.132	0.207	***	0.191	0.128	0.224	***
	(0.386)	(0.339)	(0.406)		(0.393)	(0.334)	(0.417)	
Asthma	0.064	0.054	0.068	*	0.071	0.075	0.069	
	(0.244)	(0.227)	(0.253)		(0.257)	(0.263)	(0.253)	
Allergy	0.035	0.033	0.036		0.039	0.047	0.035	
	(0.183)	(0.178)	(0.185)		(0.194)	(0.211)	(0.184)	
Age	6.188	6.169	6.198	***	6.198	6.189	6.203	
	(0.303)	(0.283)	(0.312)		(0.324)	(0.304)	(0.334)	
Female	0.483	0.482	0.483		0.503	0.502	0.503	
	(0.500)	(0.500)	(0.500)		(0.500)	(0.501)	(0.500)	
Preschool 3+	0.796	0.871	0.757	***	0.799	0.907	0.744	***
Years	(0.403)	(0.336)	(0.429)		(0.401)	(0.292)	(0.437)	
Single child	0.209	0.193	0.217	*	0	0	0	
	(0.406)	(0.395)	(0.412)		-	-	-	
3+ children	0.288	0.278	0.293		0.445	0.433	0.451	
	(0.453)	(0.448)	(0.455)		(0.497)	(0.496)	(0.498)	
First child	0.504	0.510	0.501		0.377	0.371	0.380	
	(0.500)	(0.500)	(0.500)		(0.485)	(0.484)	(0.486)	
Second child	0.347	0.359	0.341		0.4403	0.4704	0.4249	
	(0.476)	(0.480)	(0.474)		(0.4967)	(0.4999)	(0.4947)	
Third child	0.1053	0.104	0.106		0.1246	0.1308	0.1214	
	(0.307)	(0.305)	(0.308)		(0.3304)	(0.3378)	(0.3269)	
Lives with both	0.825	0.892	0.790	***	0.8933	0.9564	0.861	***
Parents	(0.380)	(0.310)	(0.407)		(0.3088)	(0.2046)	(0.3462)	
Parent full time	0.804	0.897	0.756	***	0.8215	0.9128	0.7748	***
Employed	(0.397)	(0.305)	(0.429)		(0.3831)	(0.2826)	(0.4181)	
Age of mother	36.010	38.261	34.855	***	35.895	38.377	34.623	***
	(5.166)	(4.419)	(5.142)		(4.794)	(4.013)	(4.665)	
Turkish origin	0.080	0.040	0.101	***	0.1109	0.0498	0.1422	***
	(0.272)	(0.195)	(0.301)		(0.3141)	(0.218)	(0.3495)	
Eastern European	0.120	0.084	0.138	***	0.0834	0.053	0.099	**
Origin	(0.325)	(0.278)	(0.345)		(0.2767)	(0.2243)	(0.299)	
Other foreigners	0.043	0.021	0.055	***	0.0338	0.0031	0.0495	***
	(0.203)	(0.143)	(0.227)		(0.1808)	(0.0558)	(0.2171)	
Observations	4,245	1,439	2,806		947	321	626	

Standard deviations in parentheses

Differences by parental education group are significant at 10%; ** at 5%; *** at 1%

Table 2: Linear regression models for CPM score

	Pooled Sample (1)	Parents college (2)	Parents no college (3)
Panel A: OLS (Full sample)			
Overweight	-0.316***	-0.113	-0.300***
	(0.093)	(0.172)	(0.108)
Underweight	-0.069	0.067	-0.197
	(0.169)	(0.212)	(0.236)
Low birth weight	-0.779***	-1.066***	-0.599***
	(0.178)	(0.354)	(0.202)
Ear condition	-0.167	-0.078	-0.252*
	(0.103)	(0.159)	(0.133)
Eye condition	-0.230***	-0.238*	-0.221**
	(0.085)	(0.144)	(0.105)
Mental condition	-1.067***	-0.941***	-1.047***
	(0.104)	(0.179)	(0.125)
Asthma	-0.493***	0.083	-0.699***
	(0.164)	(0.194)	(0.209)
Allergy	0.079	0.019	0.121
	(0.155)	(0.245)	(0.197)
Observations	4,245	1,439	2,806
R-squared	0.12	0.11	0.11
Panel B: OLS (Sibling sample)			
Overweight	-0.367	0.299	-0.459*
	(0.232)	(0.334)	(0.271)
Underweight	-0.087	0.279	-0.383
	(0.367)	(0.441)	(0.500)
Low birth weight	-0.841***	-0.924*	-0.620**
	(0.262)	(0.509)	(0.287)
Ear condition	-0.062	-0.395	-0.035
	(0.200)	(0.256)	(0.298)
Eye condition	-0.003	-0.024	0.057
	(0.184)	(0.274)	(0.228)
Mental condition	-1.577***	-1.065***	-1.703***
	(0.219)	(0.317)	(0.270)
Asthma	-0.279	-0.182	-0.263
	(0.264)	(0.298)	(0.371)
Allergy	-0.205	-0.259	-0.325
	(0.284)	(0.352)	(0.394)
Observations	947	321	626
R-squared	0.20	0.29	0.18

Parentheses show robust standard errors, clustered at household level.
In panels A and B, all demographic and socioeconomic characteristics listed in Table 1 are included, but not shown.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2 (continued): Linear regression models for CPM score

	Pooled sample (1)	Parents college (2)	Parents no college (3)
Panel C: Sibling fixed effects (Sibling sample)			
Overweight	0.067	0.430	0.067
	(0.383)	(0.563)	(0.437)
Underweight	-0.195	0.616	-0.813
	(0.405)	(0.579)	(0.585)
Low birth weight	-0.306	-0.052	0.484
	(0.408)	(0.652)	(0.463)
Ear condition	0.166	0.236	0.253
	(0.251)	(0.313)	(0.395)
Eye condition	0.173	-0.052	0.237
	(0.184)	(0.335)	(0.244)
Mental condition	-1.195***	-0.996***	-1.140***
	(0.287)	(0.342)	(0.393)
Asthma	-0.176	-0.088	-0.825*
	(0.380)	(0.467)	(0.475)
Allergy	0.015	-0.313	-0.201
	(0.455)	(0.375)	(0.466)
Observations	947	321	626
R-squared: Within	0.11	0.15	0.13
Between	0.03	0.04	0.01
Overall	0.05	0.05	0.02

Parentheses show robust standard errors, clustered at household level.

In panel C, coefficients for "age", "female", "preschool visit 3+ years", "first child", "second child", "third child", "lives with both parents", and "parent full time employed" are included, but not shown.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Linear probability models for verbal ability

	Pooled sample (1)	Parents college (2)	Parents no college (3)
Panel A: OLS (Full sample)			
Overweight	-0.023	0.032	-0.031
	(0.017)	(0.031)	(0.021)
Underweight	0.025	0.032	0.017
	(0.026)	(0.031)	(0.036)
Low birth weight	-0.056**	-0.047	-0.051
	(0.028)	(0.044)	(0.035)
Ear condition	-0.025	-0.020	-0.032
	(0.020)	(0.028)	(0.027)
Eye condition	-0.007	-0.003	-0.009
	(0.015)	(0.022)	(0.020)
Mental condition	-0.017	0.004	-0.018
	(0.016)	(0.025)	(0.020)
Asthma	-0.086***	-0.043	-0.098***
	(0.027)	(0.043)	(0.034)
Allergy	0.003	-0.074	0.039
	(0.032)	(0.057)	(0.038)
Observations	4,234	1,436	2,798
R-squared	0.06	0.07	0.05
Panel B: OLS (Sibling sample)			
Overweight	-0.050	0.026	-0.067
	(0.040)	(0.082)	(0.048)
Underweight	0.035	0.063	0.007
	(0.059)	(0.070)	(0.083)
Low birth weight	-0.108**	-0.061	-0.085
	(0.046)	(0.069)	(0.059)
Ear condition	-0.056	-0.088	-0.042
	(0.044)	(0.069)	(0.061)
Eye condition	-0.033	0.013	-0.045
	(0.031)	(0.042)	(0.042)
Mental condition	-0.067*	-0.040	-0.066
	(0.037)	(0.056)	(0.045)
Asthma	-0.086	-0.051	-0.098
	(0.057)	(0.083)	(0.073)
Allergy	-0.041	-0.194*	0.055
	(0.065)	(0.110)	(0.082)
Observations	944	320	624
R-squared	0.10	0.13	0.10

Parentheses show robust standard errors, clustered at household level.
In panels A and B, all demographic and socioeconomic characteristics listed in Table 1 are included, but not shown.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3 (continued): Linear probability models for verbal ability

	Pooled sample (1)	Parents college (2)	Parents no college (3)
Panel C: Sibling fixed effects (Sibling sample)			
Overweight	-0.038	0.055	-0.065
	(0.062)	(0.143)	(0.073)
Underweight	0.110	-0.025	0.082
	(0.093)	(0.109)	(0.136)
Low birth weight	-0.021	0.072	0.021
	(0.074)	(0.107)	(0.089)
Ear condition	-0.039	-0.061	-0.072
	(0.058)	(0.081)	(0.083)
Eye condition	-0.028	0.053	-0.073
	(0.039)	(0.059)	(0.055)
Mental condition	-0.097**	-0.017	-0.100*
	(0.049)	(0.094)	(0.059)
Asthma	-0.094	-0.118	-0.120
	(0.078)	(0.118)	(0.098)
Allergy	0.034	-0.011	-0.046
	(0.091)	(0.217)	(0.092)
Observations	944	320	624
R-squared: Within	0.05	0.12	0.08
Between	0.06	0.01	0.03
Overall	0.06	0.01	0.04

Parentheses show robust standard errors, clustered at household level.
In panel C, coefficients for "age", "female", "preschool visit 3+ years", "first child", "second child", "third child", "lives with both parents", and "parent full time employed" are included, but not shown.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Decomposition Analysis

	Full sample OLS	Sibling sample OLS	Sibling sample Fixed effects
Panel A: CPM score			
Mean: parents college	11.367	11.550	11.550
Mean: parents no college	10.596	10.493	10.493
Total difference	0.771	1.057	1.057
	[100.0%]	[100.0%]	[100.0%]
Prevalence effect	0.129	0.201	0.114
	(0.021)	(0.064)	(0.053)
	[16.7%]	[19.0%]	[10.8%]
Severity effect	0.094	0.167	0.080
	(0.076)	(0.146)	(0.225)
	[12.2%]	[15.8%]	[7.6%]
Total health effect	0.223	0.368	0.194
	(0.082)	(0.160)	(0.235)
	[28.9%]	[34.8%]	[18.4%]
Panel B: Verbal ability			
Mean: parents college	0.884	0.871	0.871
Mean: parents no college	0.787	0.763	0.763
Total difference	0.097	0.108	0.108
	[100.0%]	[100.0%]	[100.0%]
Prevalence effect	0.005	0.012	0.013
	(0.002)	(0.008)	(0.009)
	[5.2%]	[11.1%]	[12.0%]
Severity effect	0.017	0.021	0.057
	(0.013)	(0.029)	(0.042)
	[17.5%]	[19.5%]	[52.8%]
Total health effect	0.022	0.033	0.070
	(0.013)	(0.030)	(0.043)
	[22.7%]	[30.6%]	[64.8%]

Parentheses include standard errors. Square brackets include share of total difference in average achievement by parental education, which can be explained by health effects.

Figure 1: Cumulative frequency distribution of CPM score by parental education
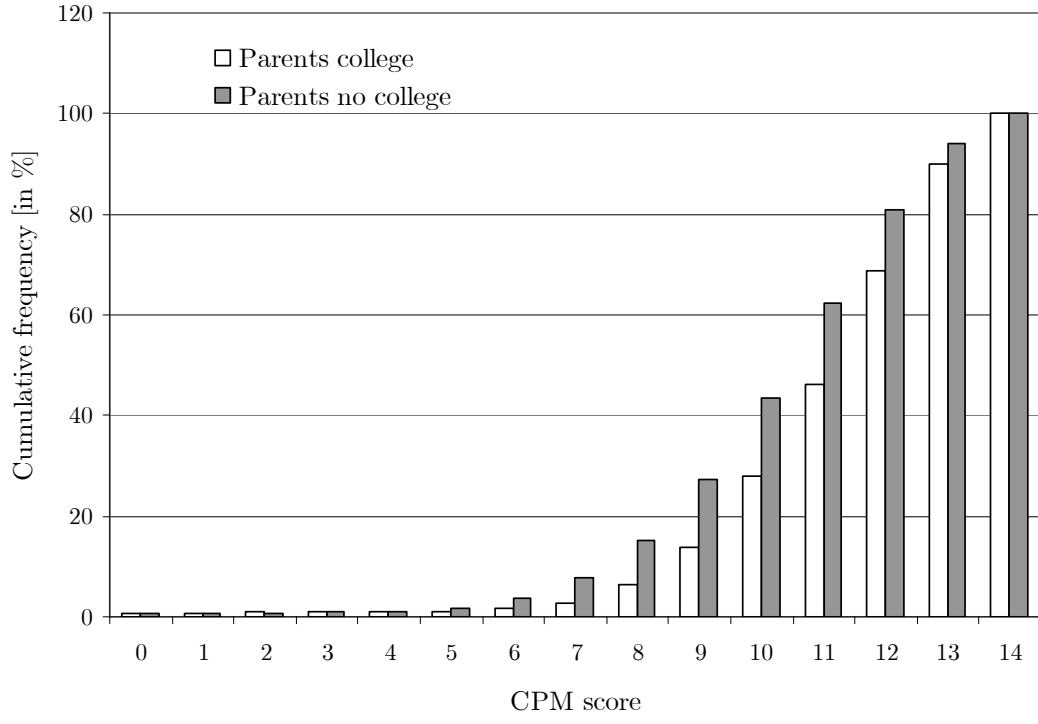
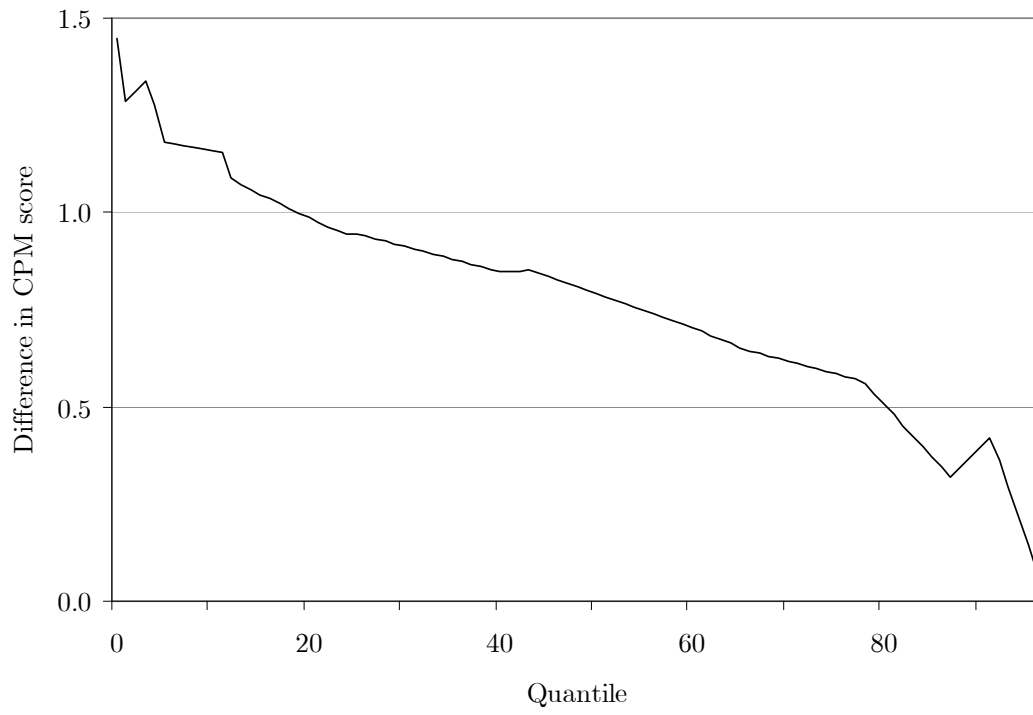Figure 2: Quantile distribution of total difference in CPM score

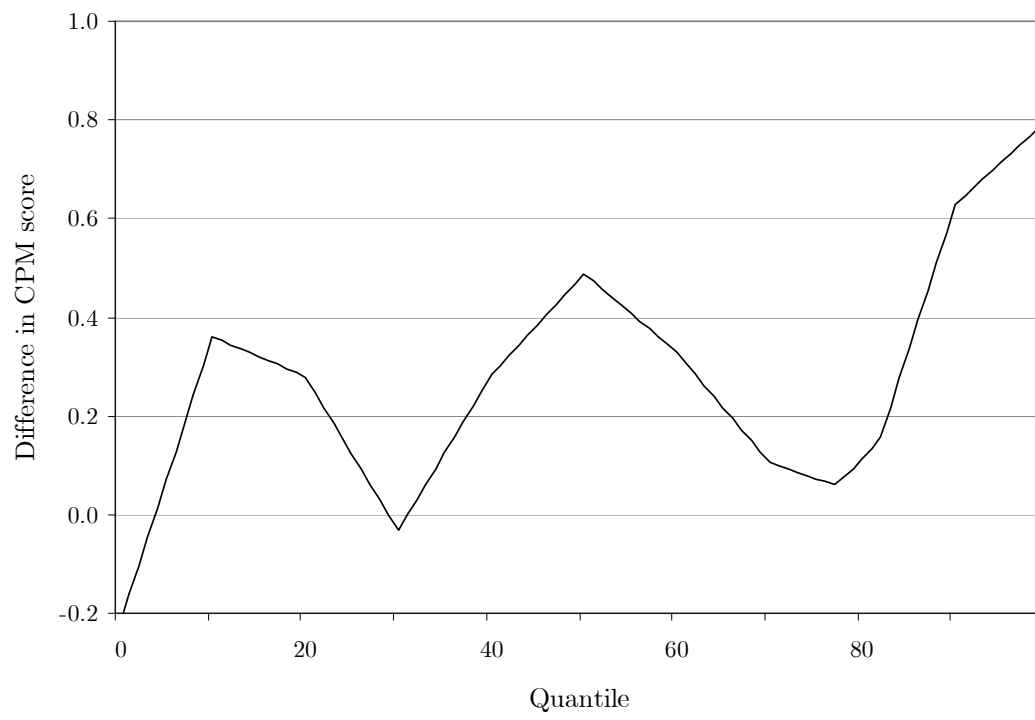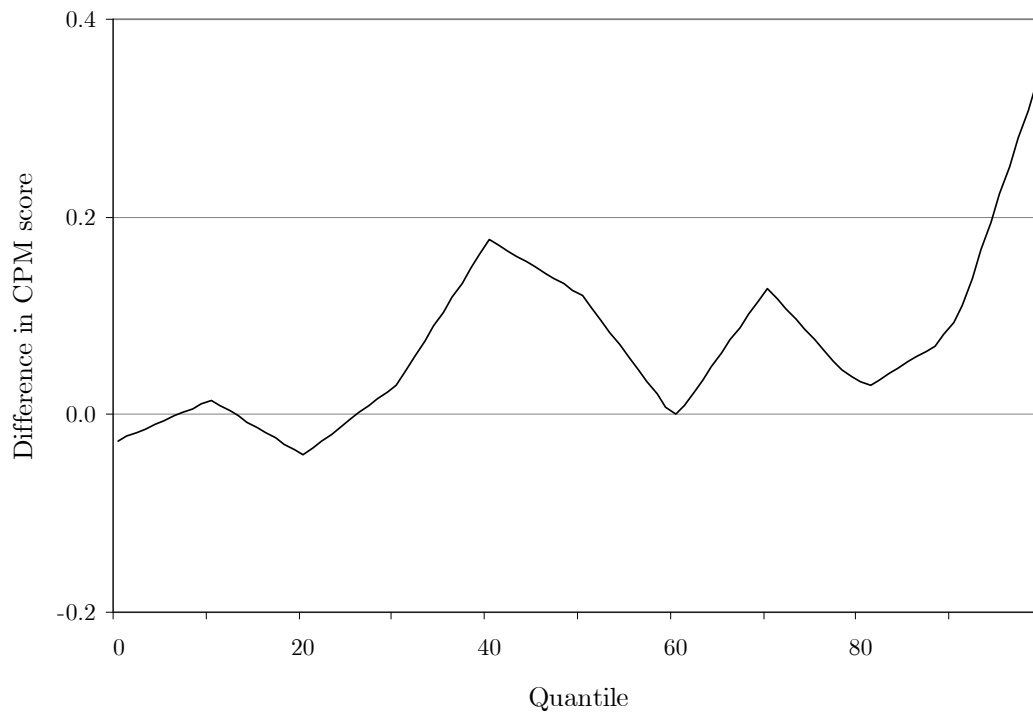Figure 3: Quantile distribution of prevalence effect

Figure 4: Quantile distribution of severity effect

Figure 5: Quantile distribution of residual effect