Hospital contacts amongst high achieving adolescents from disadvantaged socio-economic backgrounds.

John Jerrim[1]

Abstract

High achieving children from disadvantaged backgrounds are an important group for promoting social mobility. Prior research has found their educational achievement does not keep pace with their more socio-economically advantaged peers during secondary school, with early adolescence identified as a key period when their academic potential starts to go unfulfilled. This paper presents new evidence on the health outcomes of this group during this period, focusing on the contacts they make with hospital services due to mental health issues or risky health behaviours. We find high achieving children from disadvantaged backgrounds are hospitalised due to mental health issues and associated risky behaviours at a rate of around 80 per 1,000 children between the ages of 11 and 20. This is significantly above the rate of around 40 per 1,000 children for their equally able but more socio-economically advantaged peers. Our results hence provide further evidence that early adolescence is a key period during which high achieving children from disadvantaged backgrounds require further support. They may also suggest that mental health issues could be an important factor explaining in the poorer academic outcomes for high achieving disadvantaged young people.

Key Words: High-achieving disadvantaged children; risky health behaviours; mental health; social mobility.

Acknowledgments: The Nuffield Foundation is an independent charitable trust with a mission to advance social wellbeing. It funds research that informs social policy, primarily in Education, Welfare, and Justice. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org.

We would like to acknowledge the contribution of the wider ECHILD Database support and programme management.

The ECHILD Database uses data from the Department for Education (DfE). The DfE does not accept responsibility for any inferences or conclusions derived by the authors. This work uses data provided by patients and collected by the National Health Service as part of their care and support. Source data can also be accessed by researchers by applying to NHS England. We are grateful to the Office for National Statistics (ONS) for providing the Trusted Research Environment for the ECHILD Database. ONS agrees that the figures and descriptions of results in the attached document may be published. This does not imply ONS' acceptance of the validity of the methods used to obtain these figures, or of any analysis of the results. The views in this publication do not necessarily reflect the views of UCL

[1] Social Research Institute, University College London, 20 Bedford Way London, WC1H 0AL. E-mail: j.jerrim@ucl.ac.uk (John Jerrim). +44 7590 761 755

1. Introduction

There has long been interest – both in England and internationally – about the life prospects of young people from disadvantaged socio-economic backgrounds. A wide body of research has shown how this group falls behind children from more advantaged socio-economic backgrounds long before starting school (Feinstein, 2003), leading to worse outcomes in later life. This has resulted in very few young people from poor backgrounds entering professional jobs, with economic prosperity and social mobility hampered as a result (Sutton Trust, 2010). Indeed, recent research has illustrated how socio-economic gaps in educational outcomes increase markedly in England as young people progress through school (Farquharson et al., 2022).

Cunha et al. (2006) developed the theory of lifecycle skill formation in light of such observations, noting how there are key points during which children's abilities (both cognitive and non-cognitive) rapidly develop. A critical part of this theory is that – if young people start to fall behind their peers academically (as is unfortunately the case for many children from disadvantaged socio-economic backgrounds) – then it is increasingly difficult for them to catch-up. This has consequently led to great academic and public policy interest (and investment) in the early years, in order to give children from disadvantaged backgrounds the best start in life (Duncan et al., 2022). The hope is that such investment will reduce the number of disadvantaged children falling behind.

Yet the theory of life-cycle skill formation also makes another group of particular interest in terms of social mobility – initially high-achieving children from disadvantaged socioeconomic backgrounds. These are young people who have managed to overcome poverty during their early life to outperform many of their more advantaged peers during primary education. They have thus developed the platform needed to excel during secondary education, break through the glass ceiling, and eventually progress up the socioeconomic ladder, enjoying all the financial, social and health benefits this entails. Indeed, high-achieving children from disadvantaged backgrounds are perhaps the "lowest hanging fruit" in terms of promoting social mobility; they are a group of young people with the clearest potential to flourish at a top university and gain employment in a high-flying job. Adolescence is however the next tricky hurdle they face, a time during which they will experience rapid physical, social and emotional change. Unfortunately, this is a period when many disadvantaged high-achieving children may not fulfil their early potential, at least not to the same extent as their more advantaged peers (Jerrim & Carvajal, 2024).

A modest literature has emerged investigating the development of this group, mainly focused on how their educational outcomes compare to equally able children from more advantaged socioeconomic backgrounds. Jerrim & Vignoles (2013) discussed the methodological challenges with drawing comparisons across these groups, noting how estimates may be impacted by regression to the mean. This problem is caused by measured error in the test score used to separate children into low and high-achievement groups, which can lead to spurious differences in future outcomes emerging across socio-economic groups, even when no differences really exist. They go on to report that – once this methodological issue has been addressed – 3-year-olds from disadvantaged backgrounds with high cognitive test scores have similar levels of academic achievement at age 7 as their equally able but more socioeconomically advantaged peers. This is consistent with the work of Jerrim & Carvajal (2024) who found that, after addressing for the methodological issue of Kelley's paradox (a form of regression to the mean), initially high-achieving 5-year-olds from poor homes obtained similar test scores at age 11 to equally high-achieving 5-year-olds from more affluent families. These authors then show how these groups then diverge in terms of a range of social, health and educational outcomes between the ages of 11 and 17. This is consistent with the work of Crawford et al. (2017), who found between the ages of 11 and 14 to be a key period, when a notable divergence in educational outcomes between high-achieving children from different socioeconomic backgrounds emerge. Holt-White & Cullinane (2023) built upon this work, finding a 22-percentage point difference amongst high-achieving 11-year-olds from socio-economically advantaged and disadvantaged families in the probability of achieving top grades in England's high-stakes GCSE examinations. A related international literature has also emerged into so-called "resilience" (disadvantaged pupils that succeed academically). For instance, based upon the PISA 2018 study, Thomson (2021) notes how resilient students are disproportionately likely to set goals, study in more disciplined classrooms and are more likely to have a "growth mindset" than non-resilient students. The OECD (2019) use the same data to investigate the association between academic resilience and wellbeing, finding "no significant difference in well-being between academically resilient students and students who were not academically resilient" (OECD, 2019: 77).

It is notable how most previous research has focused on the progress made by high-achieving disadvantaged children while at school. The evidence base is much thinner in terms of the broader economic, health and social outcomes experienced by this group. Yet it is well-known that health issues can influence these educational outcomes (and vice-versa), with engagement in risky health behaviours – and associated mental health issues – during adolescence potentially related to why many disadvantaged high achievers start to fall behind academically. This study therefore builds upon previous research by investigating the prevalence of serious health outcomes (driven by risky behaviours or mental health issues) amongst high-achieving children from disadvantaged backgrounds during the teenage years. Previous studies into the health outcomes of this group have relied upon small scale survey data, focused on less severe outcomes, and typically included less than 300 disadvantaged high-achieving children in the sample (Jerrim & Carvajal, 2024). We, on the other hand, draw upon population level data from across three school cohorts, encompassing more than 42,000 high achieving children from disadvantaged socioeconomic backgrounds. This is important – and allows us to extend the existing evidence base – by demonstrating how the relatively mild issues recorded in previous research (e.g. higher levels of alcohol consumption, sexual activity, mild mental health symptoms) translate into more serious – and potentially life-changing – consequences.

Two research questions are addressed. We begin by documenting the prevalence of contacts with hospital services due to risky health behaviours – and associated mental health issues – across socioeconomic groups. This provides an important first descriptive account of how contacts with hospital services vary across young people from different socioeconomic backgrounds during adolescence:

- Research question 1. How do contacts with hospital services due to mental health issues and risky health behaviours vary across socioeconomic groups? How does this change during adolescence?

This is then followed by consideration of how children's socioeconomic background intersects with their academic achievement. Our particular focus is the prevalence of hospital contacts amongst high-achieving children from disadvantaged social backgrounds, and how this compares to their more advantaged peers.

- Research question 2. How do contacts with hospital services due to mental health issues and risky health behaviours differ between high-achieving children from different socioeconomic backgrounds? How does this change during adolescence?

The paper now proceeds as follows. In section 2 we describe the administrative data we analyse, along with our methodological approach. Results are reported in section 3, followed by conclusions in section 4.

2. Data and methodology

Overview

The data we use are drawn from the Education and Child Health Insights from Linked Data (ECHILD); a resource that combines information from England's National Pupil Database (NPD) to records from the Health Episode Statistics (HES)[2]. The latter includes information from both inpatient and outpatient services, with data available (at the time of writing) up until 2021. The match quality is high, with over 90% of children within the NPD being successfully match to their hospital records. These data include information on all contacts young people had with hospital care providers throughout childhood, though our focus is on those that occurred after young people entered secondary school. Ethical approval was obtained from the UCL Institute of Education ethics committee to access these data from the Office for National Statistics (ONS) Secure Research Service (SRS) to conduct this research (REC1821).

Our analysis focuses on three adjacent school cohorts – children born between September 2000 and August 2001 (cohort A), between September 2001 and August 2002 (cohort B) and those born between September 2002 and August 2003 (cohort C). The use of three school cohorts has the advantage of increasing the sample size. This is important for our second research question (e.g. investigating hospital contacts amongst high-achieving children from disadvantaged socio-economic backgrounds), particularly since some of our outcome measures – receipt of hospital care related to specific diagnoses – are relatively rare. Young people in these cohorts entered secondary school between September 2012 and 2014, took their GCSES between 2017 and 2019, and then entered university (if they chose to do so) between 2019 and 2021. Information is available on hospital contacts made by these cohorts until they are 19 (cohort C), 20 (cohort B) or 21 (cohort A).

Contact with hospital services

For each contact young people had with hospitals, International Classification of Diseases (ICD-10) codes are available. These capture the diagnoses made by staff at the hospital to classify the contact into different groups. It is possible for multiple ICD-10 codes to be assigned to a hospital contact if it falls into more than one category; for instance, a hospital admission due to self-injury as a primary diagnosis, but with depression also identified as a secondary cause. This information is typically recorded at the three-digit level. Burns et al. (2012, p.138)

[2] See Ramzan et al. (2023: Appendix 2) for further details on the linkage procedures used.

discuss the accuracy of ICD-10 codes in British administrative data, concluding that "*routinely collected data are sufficiently robust to support their use for research*". As these data capture contacts with hospitals, our focus is upon severe health outcomes. A limitation of these data is that we are unable to observe less severe health issues, such as contacts made with primary health care provides (e.g. prescriptions made by GPs), or undiagnosed/untreated health issues.

We broadly follow the Child and Adolescent Mental Health Disorders Classification System - (CAMHD-CS) to identify hospital contacts related to child mental health disorders (Zima et al., 2020). This takes the ICD-10 codes and classifies contacts with healthcare services into one of thirty mental health disorder groups, such as anxiety disorders, self-injury and substance abuse. We focus on a sub-set of these categories, combining some together to increase the available cell sizes. Our analysis thus focuses on hospital contacts due to:

(a) Alcohol or drug misuse
(b) Eating disorders
(c) Self-harm
(d) Personality disorders (e.g. bipolar, conduct disorders, obsessive-compulsive disorder)
(e) Mental health issues (e.g. anxiety, depression)
(f) Any of the above plus other related miscellaneous diagnoses (e.g. sleeping disorders)

The full set of ICD-10 codes used to define each outcome can be found in Appendix I. For females, we are additionally able to observe whether they made contact with a hospital due to pregnancy. This, on most occasions, records successful childbirth. As we are unable to observe terminations, this is largely an indicator of a *completed* pregnancies (i.e. giving birth to a child) rather than of falling pregnant per se.

Measurement of academic achievement

Our primary measure of children's academic achievement is their performance in England's Key Stage 2 Statutory Assessment Tests (SATs). These are national examinations taken at the end of primary school (age 10/11) and play a prominent role in school accountability (Jerrim, 2021). As part of the SATs, young people sit around four hours of test material over a four-day period, with this designed to capture their abilities in English and mathematics. A limitation of the SATs is that they are typically only taken by pupils within state schools, and not those being educated within the independent sector. Within our analysis we first standardise young people's English and mathematics scores to mean zero and standard deviation one within their school cohort, with the average across the two subjects then taken.

Measurement of socioeconomic status

We draw upon two pieces of information to construct a socio-economic status index. The first is information on children's eligibility for Free School Meals (FSM). This is an income-based benefit that is widely used to identify young people living in low-income households, with various studies illustrating its validity as a marker of socio-economic disadvantage (Hobbs & Vignoles, 2010; Taylor, 2018; Jerrim, 2023). We can observe this measure each year that a child is enrolled in school. The second piece of information is the Income Deprivation Affecting Children Index (IDACI) decile based upon the child's home postcode. This captures the proportion of children living in income-deprived families within the small geographic area (around 1,500 people) that a child lives. We use information from IDACI at three points during childhood (recorded in years 2008, 2012 and 2015).

Following Jerrim (2023) we combine these indicators into a single continuous scale. A two-parameter item-response theory (IRT) model is estimated, with the annual FSM indicators and IDACI deciles used to create a latent indicator of a child's socio-economic position. Expected A Posteriori (EAP) scores are then created for each child, capturing the relative (dis)advantage of their socio-economic background.

Methodology

To address our first research question, we investigate the prevalence of hospital contacts due to risky health behaviours across deciles of our socioeconomic status scale. We begin by considering any hospital contact after young people have entered secondary school (from age 11 to approximately age 20), before exploring how these contacts vary by school year group. Analogous estimates for differences in hospital contacts across deciles of Key Stage 2 scores are provided in Appendix A and Appendix B.

To address our second research question, we will consider how the prevalence of hospital contacts compare across high-achieving children from different socioeconomic backgrounds. We define "high achieving" as being in the top quartile of the Key Stage 2 test score distribution, with socioeconomic advantage / disadvantage defined as the top/bottom quartile of the socioeconomic status scale. Our decision to use quartiles – rather than deciles as in research question 1 – is for consistency with previous literature (e.g. Jerrim & Vignoles, 2013) and to ensure sufficient cell sizes in our analyses (given that some of the outcomes we investigate – particularly when presenting results by school year – are relatively rare).

Our analysis begins by documenting how hospitalisations for our diagnoses of interest vary across 16 socio-economic-by-achievement quartiles. However, as noted by Jerrim & Vignoles (2013) and Jerrim & Carvajal (2024), these estimates may lead one to overstate the difference in hospital contacts between high-achieving children from different socioeconomic backgrounds due to "Kelley's paradox" (Wainer & Brown, 2007) – a form of regression to the mean. In essence, the measurement error in the Key Stage 2 scores used to assign pupils into the "high achieving" group will result in some misclassification (i.e. some individuals will be classified as "high achievers" when they are not). As the rate of misclassification will differ by socio-economic background, a statistical artifact may emerge, whereby high achieving disadvantaged children may falsely appear to suffer worse outcomes than their equally high achieving but more socioeconomically advantaged peers (even when no difference really exists).

We apply the four-step procedure proposed by Jerrim & Carvajal (2024) to test the sensitivity of results to this issue. In the first step, we estimate the raw difference in hospital contacts between children with test scores in the top Key Stage 2 quartile ("high achievers") amongst those from the top and bottom socioeconomic status quartile (children from "advantaged" and "disadvantaged" backgrounds). These results are essentially equivalent to assuming the Key Stage 2 test captures 11-year-olds academic skills with perfect reliability.

In step 2, we estimate the difference in the academic abilities of high achieving children from different socioeconomic backgrounds under different assumptions of Key Stage 2 reliability, using the formula presented in Wainer & Brown (2007):

$$\tau_i = \rho(x_i) + (1 - \rho).u_g \quad (1)$$

Where:

τ_i = Child i's academic abilities at the end of primary school.

ρ = The reliability of the Key Stage 2 test.

x_i = Child i's score on the Key Stage 2 test.

u_g = The average Key Stage 2 test score of the socioeconomic group (g) to which the child belongs (e.g. the average Key Stage 2 score of children from disadvantaged backgrounds).

These estimates are produced under three different assumptions of Key Stage 2 test reliability (ρ): 1 (perfect reliability), 0.7 (modest levels of test reliability) and 0.5 (low levels of test reliability). In step 3, we estimate the strength of the association between young people's academic abilities at the end of primary school and the chances of making a relevant hospital contact. Finally, we take the "raw" difference from step 1 and adjust it downwards by the product of the estimated difference in high achieving advantaged and disadvantaged children's academic abilities (from step 2) with how strongly this difference in abilities is linked to hospital contacts (from step 3). This approach allows us to present estimates recognising Key Stage 2 tests are an imperfect measure of children's academic abilities (ρ = 0.7), accompanied by a set of upper (ρ = 1) and lower (ρ = 0.5) bounds. See Jerrim & Carvajal (2024: section 3) for further details about this approach. As we will illustrate in the following section, most of our substantive results remain unchanged whether we correct estimates for Kelley's paradox or not.

3. Results

Research question 1. The socio-economic gap in hospital contacts related to mental health issues and risky health behaviours

Figure 1 begins by plotting the rate of hospital contacts per 1,000 children by socio-economic status decile. The table at the bottom presents results for all outcomes, with a selection of these plotted in the figure above. Red shading in the table should be read vertically, with darker colours indicating more frequent contacts with hospital services.

<< Figure 1 >>

On the whole, the socio-economic gradient in hospitalisations is relatively shallow between the bottom decile (the most advantaged 10% of children) and the 7th decile. For instance, the overall rate for any hospital contact stemming from mental health issues or risky health behaviours is 46 per 1,000 children for those from the most advantaged socio-economic decile, compared to 56 per 1,000 children amongst those in the 7th socio-economic decile. However, there is then a sizeable increase in the hospital contact rate amongst the most disadvantaged 30% of the population. Take hospital contacts related to alcohol or drug misuse; the rate almost doubles for young people in the 8th socio-economic decile compared to the 7th (≈33 per 1,000 compared to ≈18 per 1,000). This then increases still further up to ≈47 per 1,000 children amongst the most disadvantaged socioeconomic group.

There is a particularly noticeable socio-economic gradient with respect to hospital contacts related to pregnancy (which largely reflects differences in child birth). Amongst the most advantaged 10% of young people, only around 9 in every 1,000 young women will have made contact with a hospital due to pregnancy by age 20. In comparison, the rate is almost three times higher (around 24 in every 1,000 young women) for those from an "average" socioeconomic background. This then spikes up to 66 in every 1,000 young women within the 8th socioeconomic status decile, and over 100 amongst the most disadvantaged group (over 10 times more than the least advantaged group). This demonstrates how there are substantial socioeconomic differences in the probability of teenage parenthood in England.

A similar – though less marked – increase according to socio-economic status can be observed for most of the other outcomes considered. The only notable exception is eating disorders where – if anything – the prevalence of hospital contacts is slightly higher amongst young people from advantaged socio-economic backgrounds. This finding is consistent with previous research, which has found eating disorders to be more common amongst higher-achieving and higher socio-economic status groups (Weckstrom et al., 2023).

There are two potential explanations for why we observe there to be a spike in hospital contacts amongst the 8th to 10th socio-economic decile (in comparison to the relatively shallow gradient observed between the 1st and 7th socioeconomic deciles). The first is that there are important non-linearities in hospital contacts for risky health behaviours and mental health issues by socio-economic status. That is, while most of the population rarely make contact with hospitals for such reasons, the most disadvantaged 30% of society have to make use of these services much more frequently. The second explanation is that this may in part be a reflection of the measurement properties of our socio-economic status scale. As FSM eligibility primarily captures socio-economic disadvantage, then the spike observed between the 7th and 8th decile could be due to increased measurement precision at this point in the socio-economic distribution.

Figure 2 builds upon these findings by demonstrating how hospital contacts for four of our outcomes change across socio-economic decile by school year group. Each line refers to a different socio-economic decile, with school year group running along the horizontal axis, and the hospital contact rate on the vertical axis.

<< Figure 2 >>

Starting with panels (a) - contact with hospitals for any risky behaviour or mental health issue - and (b) -contacts related to alcohol or drug misuse - a socio-economic difference can be observed even in the first year of secondary school (Year 7). While the absolute rate is very small, it is still around three times greater for those from the most disadvantaged 10% of children compared to the most advantaged decile (3.5 per 1,000 compared to 1.1 per 1,000). This gap in the hospital contact rate then grows throughout secondary school. The increase in hospitalisations amongst those from disadvantaged backgrounds is particularly steep during Year 9 (age 13/14) and Year 10 (age 14/15). For instance, the overall hospital contact rate amongst the most disadvantaged decile almost triples from 6.4 per 1,000 in Year 8 to 16.9 per 1,000 in Year 10. While there is also an increase amongst more advantaged socio-economic groups over the same period, it is much more muted. Indeed, a Year 8 pupil from the most disadvantaged 10% of families has roughly the same hospital contact rate related to risky health behaviours as a Year 10 pupil from an average socioeconomic background. This holds true both in terms of hospital contacts for any of our outcomes (panel a) or when looking specifically at those related to alcohol or drug misuse (panel b). The other notable spike in these graphs occurs during Year 13 – as young people turn age 18 – where the socio-economic gap in hospital contacts increases still further.

Panel (c) presents a slightly different pattern in terms of self-harm. Here there is a clear jump in hospital contacts during Year 9 and Year 10 across all socio-economic groups, though the increase is particularly marked for those from the poorest backgrounds. Indeed, for the bottom socio-economic group, the rate triples from 3.6 per 1,000 in Year 8 to 10.2 per 1,000 in Year 10. However, from Year 11 onwards, the rate then starts to fall amongst the most disadvantaged groups, while staying relatively stable for those from "average" and more advantaged socio-economic backgrounds. Together, the results point towards Year 9 and Year 10 as a particularly risky period for young people from disadvantaged social backgrounds with respect to serious injury related to self-harm.

To conclude, panel (d) focuses on hospital contacts related to pregnancies. It is immediately clear that young women from disadvantaged backgrounds give birth earlier, with the socio-economic gap widening appreciably between the end of secondary school (Year 11) and age 20. Indeed, one in every 200 young women from a disadvantaged background will make contact with a hospital due to a pregnancy while they are in Year 11 (and sitting GCSE examinations). This rate then increases exponentially amongst the most disadvantaged group as soon as compulsory education (Year 11) has finished, reaching 14 per 1,000 children during

Year 12 (age 16/17) and 25 per 1,000 during Year 13 (age 17/18). While the rate also increases for other socio-economic groups, the trajectory is not nearly as sharp. Note that, as these contacts are mostly related to child births, the results presented in panel (d) reflect a combination of factors. This potentially includes disadvantaged young women being sexually active earlier, being less likely to use effective contraception and being more likely to make a conscious decision to become a young mother (Singh et al., 2001; Aluga & Oklie, 2021).

Research question 2. Contact with hospital services for high-achieving children from different socio-economic backgrounds

In Table 1 we illustrate the rate of hospital contacts for five outcomes across 16 socio-economic-by-achievement groups. Note that these estimates do not attempt to adjust for Kelley's paradox, with it therefore prudent to interpret differences across groups as upper bounds.

<< Table 1 >>

Panel (a) presents the rate of hospital contacts that occurred due to any of our outcomes of interest by age 20. Even amongst high-achieving children there is a clear socio-economic gradient. Young people from the most disadvantaged 25% of backgrounds with a Key Stage 2 score in the top quartile had an 8.6% (i.e. 86 children out of every 1,000) chance of having contact with hospital services by age 20 related to a risky health behaviour or mental health issue. This is roughly twice the rate of their peers with similar Key Stage 2 scores from the most advantaged socio-economic quartile (4.2% or 42 children in every 1,000). Indeed, the hospital contact rate amongst disadvantaged high achievers is above that for low-achieving children from the most advantaged backgrounds (where the chance is 6% or 60 in every 1,000 children).

The same pattern is replicated in panels (b) and (c) where we focus upon hospital contacts related to alcohol/drug use and self-harm. These raw figures suggest that approximately 13 in every 1,000 high-achieving high socio-economic status children will require hospital treatment for an issue related to drug/alcohol use up to age 20, with the rate for their equally able but socio-economically disadvantaged peers around 2.5 time greater (≈31 in every 1,000 children). This is again notably higher than for low-achieving children from affluent backgrounds (18 in every 1,000 children). Similar findings also emerge for self-harm.

The results for pregnancy in panel (e) are particularly striking, where – at least in these raw figures – the interplay between socio-economic background and academic achievement is particularly strong. High-achieving girls from disadvantaged backgrounds are 10 times more likely to make contact with a hospital due to pregnancy by age 20 than their peers with similar Key Stage 2 scores from the most affluent backgrounds (rates of ≈35 per 1,000 compared to ≈3 per 1,000 respectively). These figures are then dwarfed, however, by the rate for low-achieving young women from disadvantaged backgrounds (129 per 1,000) which is 40 times higher than for the high-achieving high socio-economic group.

The final point of note from Table 1 is with respect to eating disorders (panel c) where the pattern again differs to the other outcomes. Most notably, the hospital contact rate is slightly higher amongst young people with high levels of achievement, with variation across socio-economic groups less clear.

In the methodology section we noted how the estimates presented in Table 1 may be impacted by Kelley's paradox, and hence overstate the difference in hospital contacts made by high-achieving young people from different socioeconomic backgrounds. Table 2 therefore explores the sensitivity of our results to this issue, applying the approach suggested by Jerrim & Carvajal (2024). Panel (a) presents the estimated *difference* in the hospital contact rate between high-achieving children from different socio-economic backgrounds under different assumptions of test reliability (labelled "rho"). Panel (b) then converts these into estimates of the hospital contact rate for high-achieving children from disadvantaged backgrounds.

<< Table 2 >>

For most outcomes, correcting the estimates presented in Table 1 for Kelley's paradox makes little substantive difference to the results. While the difference in the hospital contact rate between high achieving children from different socio-economic backgrounds narrows slightly when test reliability is assumed to be less than perfect, the change in the estimates is relatively small. Take self-harm, for example. When assuming Key Stage 2 scores measure children's achievement with perfect reliability ($\rho = 1$) the socio-economic difference in the hospitalisation rate due to self-harm stands at 26.2 per 1,000 children (46.3 versus 20.1). If one instead assumes test reliability of 0.7 – quite a substantial amount of measurement error – the difference marginally falls to 23.5 per 1,000 (i.e. a rate of 43.7 for high achieving disadvantaged children compared to 20.1 for their equally high achieving but more socio-economically advantaged peers). Indeed, even under low levels of test reliability ($\rho = 0.5$) we estimate that the hospital contact rate due to self-harm amongst disadvantaged high achievers would be double that of those with similar abilities from the most advantaged families. Similar results hold with respect to most of the other outcomes considered.

The one notable exception is for contacts made with hospital services related to pregnancies, where the magnitude of the difference between high achievers from different socioeconomic backgrounds is more sensitive to the assumption one makes regarding test reliability. This is likely due to this outcome being more closely related to academic achievement than the other measures (as exhibited in Table 1). Our central estimate ($\rho = 0.7$) continues to indicate a sizeable difference in hospital contacts related to pregnancies between high achieving children from advantaged and disadvantaged socio-economic backgrounds (3.4 versus 20.1 per 1,000). If, however, Key Stage 2 scores are assumed to have only low levels of reliability ($\rho = 0.5$) the difference shrinks to almost zero. Our overall interpretation of this result is that, while our estimates continue to suggest there are substantial differences in hospital contacts related to pregnancy between high-achieving young women from rich and poor backgrounds, the magnitude of this difference is measured with some imprecision.

Figure 3 turns to how the hospital contact rate changes with age across four socio-economic-by-achievement groups:

- High-achieving children from advantaged socio-economic backgrounds.
- High-achieving children from disadvantaged socio-economic backgrounds.
- Low-achieving children from advantaged socio-economic backgrounds.
- Low-achieving children from disadvantaged socio-economic backgrounds.

Note that these refer to raw estimates where no adjustment has been made for Kelley's paradox (we shall return to this point below).

<< Figure 3 >>

Starting with panels a (any risky behaviour) and b (alcohol or drug misuse) the period between Year 7 and Year 10 seems key. During this time, the trajectory of hospital contacts for high achieving children from poor backgrounds (solid black line) tracks the rate of their low achieving peers (dashed grey line). In other words, the change during the early stages of secondary school seems to be more strongly associated with socio-economic status than it does to academic achievement. This leads to a gap in hospital contacts between high achieving children from advantaged (dashed black line) and disadvantaged (solid black line) socio-economic backgrounds. While this gap does not then appreciably widen after Year 10, it also does not narrow. Socio-economic inequalities in hospital contacts due to risky health behaviours amongst high achieving young people thus seem to emerge early in secondary school and are then maintained throughout adolescence. Moreover, as we illustrate in Figure 4, there is virtually no change to this result if one adjusts estimates for Kelley's paradox (see Appendix H for further details). Interestingly, low achieving low socio-economic status children seem to suffer the sharpest rise in hospital contacts after secondary school (i.e. from Year 12 and beyond).

<< Figure 4 >>

Panel (c) for self-harm exhibits an even sharper concordance between socio-economic background and hospital contacts, regardless of achievement levels. For instance, the increase between Year 7 and Year 10 is almost identical across low socio-economic status students, regardless of whether they are a higher or lower achiever at school (comparison of the solid black and grey dashed lines). There does, however, seem to be a narrowing of the gap for high achievers from different socio-economic backgrounds from Year 11 onwards, declining from a difference of ≈6 per 1,000 children in Year 10 (age 14/15) to a difference of ≈2.5 per 1,00 in Year 13 (age 17/18).

To conclude, panel (d) of Figure 4 presents the estimates for pregnancy. The exponential increase in hospital contacts due to pregnancy amongst low achieving low socio-economic status girls after Year 10 is striking. Likewise, it is also interesting to observe how the increase in the pregnancy rate for high achieving girls from disadvantaged socio-economic backgrounds is very similar to the increase (in both timing and magnitude) for high socio-economic status girls with low levels of Key Stage 2 achievement. However, as illustrated in Appendix H (and consistent with our previous discussion of Table 2) the trajectory of hospital contacts related to pregnancy for disadvantaged high achievers is measured with a degree of uncertainty (due to Kelley's paradox). We thus suggest that this finding is interpreted with care.

4. Conclusions

Socioeconomically disadvantaged children with high levels of early achievement are perhaps the lowest hanging fruit for promoting social mobility. Despite their impoverished upbringing, these children have managed to develop a strong academic platform upon which they should be able to build during secondary school. Unfortunately, previous work has shown this is not always the case (Crawford et al., 2017). England is a less equal, efficient and socially mobile country as a result.

Previous research has focused upon the future academic outcomes of this group, such as the progress they make while at school and whether they go on to attend a high-status university (Crawford et al., 2017). Yet emerging evidence highlights the broader challenges faced by disadvantaged high achievers, particularly during adolescence (Jerrim & Carvajal, 2024). This includes falling into a more troublesome peer group, brushes with the law and signs of deteriorating mental health.

The present paper has built upon this work by documenting – for the first time – how the challenges faced by disadvantaged high-achievers translate into contacts made with hospital services due to risky health behaviours and associated mental health issues. This is important as such hospital contacts are potentially a marker for broader problems that may hamper the prospects of these individuals over the longer term.

Our findings illustrate how there are sizeable socioeconomic differences in hospital contacts due to mental health issues and risky health behaviours (with the important exception of eating disorders). For instance, 47 in every 1,000 children from the most disadvantaged 10% of households will be admitted to hospital by age 20 due to an alcohol or drug related incident, compared to only 18 in every 1,000 children from an "average" background (and 15 in every 1,000 children from the most socioeconomically advantaged group). Yet, even amongst high achieving children, there continues to be sizeable differences in contacts with hospital services between those from the most and least advantaged families. For instance, after accounting for Kelley's paradox, we find that around 80 in every 1,000 disadvantaged high-achieving children makes contact with a hospital due to a mental health issue or a risky health behaviour before age 20, compared to around 40 of their equally high-achieving but more socioeconomically advantaged peers.

How do these findings compare to previous studies? The fact hospital contacts for mental health issues and risky behaviours increase amongst disadvantaged high achievers from the start of secondary school may help explain the educational trajectories documented by Crawford et al. (2017). In particular, the point when this group starts to fall behind their equally able but more advantaged peers academically coincides with when their contacts with secondary healthcare services for risky behaviours and associated mental health issues starts to increase. This is also consistent with the emerging findings of Jerrim & Carvajal (2024), with age 11 to 14 identified as a key transitional period in these individuals' lives.

The following limitations should be noted, however, when considering these results. First, our outcome measure – hospital contacts – captures extreme health outcomes. The analysis therefore does not capture milder health issues amongst disadvantaged high achievers, including those requiring treatment from primary care services (e.g. prescriptions made by General Practitioners) or problems that go undiagnosed or untreated. Future research may seek to broaden our findings to these milder outcomes, either via linkages to primary care records (something not currently possible in England) or via the collection of large-scale survey data. Second, relatedly, our analysis of hospital contacts related to pregnancies is almost entirely based upon successful childbirths. The data available does not capture information on terminations, with our analysis therefore partly reflecting conscious decisions by young people to have a child. Readers should keep this in mind when interpreting these results. Third, our measure of high achievement is based on children's performance in England's Key Stage 2 examinations. While these are lengthy assessments of 11-year-olds mathematics and English skills, we recognise they are not a perfect measure of the full spectrum of children's abilities. We have, however, illustrated how our results are robust to potential measurement error in these data, and that estimates of socioeconomic differences in high-achieving children's health outcomes are largely unaffected by Kelley's paradox. Finally, our measure of socioeconomic status has been based upon the number of years children have been eligible for free school meals and indicators of local area deprivation (IDACI). Thus, while our data are likely to capture socioeconomic disadvantage reasonably well – given this is what free school meals essentially measures (Jerrim, 2023) – children in the most advantaged socio-economic quartile may be classified with a degree of error (with few having ever been FSM eligible and thus largely classified based on IDACI). Further data linkages may again aid this aspect of the research, for instance by integrating information on household income into these data from tax records.

Despite these limitations, our findings continue to have potentially important implications. In conjunction with previous research, they point towards the broader set of challenges that high-achieving children from disadvantaged socioeconomic backgrounds encounter during early adolescence. It thus seems likely that the decline in this group's academic performance relative to their equally able but more advantaged peers is intertwined with the wider social, familial and environmental difficulties they face. Thus, while supporting this group academically at school is important, this needs to be accompanied by arrangements to help them to manage broader issues in their lives. This requires a joined-up approach across education, social care and health services. Until this happens, the full potential of this group is likely to continue to go unfulfilled.

References

Aluga, D. & Okole, E. (2021). Socioeconomic determinants of teenage pregnancy and early motherhood in the United Kingdom: A perspective. *Health Promotion Perspectives* 11(4): 426-429.

Beck, K. C., Røhr, H. L., Reme, B.-A., & Flatø, M. (2023). Distressing testing: A propensity score analysis of high-stakes exam failure and mental health. *Child Development*, 00, 1–19. https://doi.org/10.1111/cdev.13985

Bergold, S., Kasper, D., Wendt, H., & Steinmayr, R. (2020). Being bullied at school: The case of high-achieving boys. *Social Psychology of Education*, *23*, 315–338.

Bonell, C., Beaumont, E., Dodd, M., Elbourne, D. R., Bevilacqua, L., Mathiot, A., McGowan, J., Sturgess, J., Warren, E., Viner, R. M., & Allen, E. (2019). Effects of school environments on student risk-behaviors: evidence from a longitudinal study of secondary schools in England. *Journal of Epidemiology and Community Health*, *73*(6), 502-508.

Burns, E. M., Rigby, E., Mamidanna, R., Bottle, A., Aylin, P., Ziprin, P., & Faiz, O. D. (2012). Systematic review of discharge coding accuracy. *Journal of public health (Oxford, England)*, *34*(1), 138–148. https://doi.org/10.1093/pubmed/fdr054

Candito, J. (2003). Just a Stupid Kid: Low Self-Esteem and Alcoholism

Christie, D., & Viner, R. (2005). Adolescent development. *BMJ (Clinical research ed.)*, *330*(7486), 301–304. https://doi.org/10.1136/bmj.330.7486.301

Eren, F., Çete, A. Ö., Avcil, S., & Baykara, B. (2018). Emotional and Behavioral Characteristics of Gifted Children and Their Families. *Noro psikiyatri arsivi*, *55*(2), 105–112. https://doi.org/10.5152/npa.2017.12731

Fernández-García, O., Gil-Llario, M. D., Castro-Calvo, J., Morell-Mengual, V., Ballester-Arnal, R., & Estruch-García, V. (2022). Academic Perfectionism, Psychological Well-Being, and Suicidal Ideation in College Students. *International journal of environmental research and public health*, *20*(1), 85. https://doi.org/10.3390/ijerph20010085

Hobbs, G. & Vignoles, A. 2010. Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal* 36(4): 673-690.

Holm, M. E., Korhonen, J., Laine, A., Björn, P. M., & Hannula, M. S. (2020). Big-fish-little-pond effect on achievement emotions in relation to mathematics performance and gender. *International Journal of Educational Research*, 104, 101692. https://doi.org/10.1016/j.ijer.2020.101692.

Jerrim, J., & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176: 887–906.

Jerrim, J., & Sims, S. (2019). How Do Academically Selective School Systems Affect Pupils' Social-Emotional Competencies? New Evidence From the Millennium Cohort Study. American Educational Research Journal, 56(5), 1769-1799. https://doi.org/10.3102/0002831219830965

Jerrim, J. (2021). National tests and the wellbeing of primary school pupils: new evidence from the UK. *Assessment in Education: Principles, Policy & Practice*, *28*(5-6), 507-544. https://doi.org/10.1080/0969594X.2021.1929829.

Jerrim, J. (2023). Measuring parental income using administrative data. What is the best proxy available? *Research Papers in Education*.

Jerrim, J., & Carvajal, M. (2024). What happens to bright 5-year-olds from poor backgrounds? Longitudinal evidence from the Millennium Cohort Study.

Källmén, H., & Hallgren, M. (2021). Bullying at school and mental health problems among adolescents: a repeated cross-sectional study. *Child and Adolescent Psychiatry and Mental Health*, *15*(1), 74. https://doi.org/10.1186/s13034-021-00425-y.

Karpinski, R. I., Kinase Kolb, A. M., Tetreault, N. A., & Borowski, T. B. (2018). High intelligence: A risk factor for psychological and physiological overexcitabilities. *Intelligence*, *66*, 8-23. https://doi.org/10.1016/j.intell.2017.09.001.

Kiessling, L., & Norris, J. (2023). The Long-Run Effects of Peers on Mental Health. *The Economic Journal*, *133*(649), 281–322. https://doi.org/10.1093/ej/ueac039.

Loyalka, P.; Zakharov, A. & Kuzmina, Y. (2018). Catching the Big Fish in the Little Pond Effect: Evidence from 33 Countries and Regions. *Comparative Education Review* 62(4): 542-564.

Lu, B., Anders, J., Siddiqui, N., & Shao, X. (2023). How do academic selection systems affect pupils' educational attainment? New evidence from an analysis of large-scale data on England. *Educational Review*. https://doi.org/10.1080/00131911.2023.2240977.

Luthar, S. S., Kumar, N. L., & Zillmer, N. (2020). High-achieving schools connote risks for adolescents: Problems documented, processes implicated, and directions for interventions. *The American psychologist*, *75*(7), 983–995. https://doi.org/10.1037/amp0000556

Marsh, H. W., Seaton, M., Trautwein, U., et al. (2008). The Big-fish–little-pond-effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research. *Educational Psychology Review*, *20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6.

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The Reciprocal Internal/External Frame of Reference Model: An Integration of Models of Relations Between Academic Achievement and Self-Concept. *American Educational Research Journal*, 48(6), 1315-1346. https://doi.org/10.3102/0002831211419649

Murphy, R., & Weinhardt, F. (2020). Top of the Class: The Importance of Ordinal Rank. *The Review of Economic Studies*, *87*(6), 2777–2826. https://doi.org/10.1093/restud/rdaa020.

Ramzan F, Mc Grath-Lone L, Gilbert R, Blackburn R, Ruiz Nishiki M, Lilliman M, Stone T, Jay M, Harron K. (2023). Education and Child Health Insights from Linked Data (ECHILD): an introductory guide for researchers. UCL.

Rice, Frederickson, N., Shelton, K., McManus, C., Riglin, L., & Ng-Knight, T. (2015). *Identifying factors that predict successful and difficult transitions to secondary school*. The Nuffield Foundation. http://nuffieldfoundation.org/sites/default/files/files/STARS_report.pdf

Schall, J., Wallace, T. L., & Chhuon, V. (2016). 'Fitting in' in high school: how adolescent belonging is influenced by locus of control beliefs. *International Journal of Adolescence and Youth*, *21*(4), 462-475. DOI: 10.1080/02673843.2013.866148.

Singh, S.; Darroch, J. & Frost, J. (2001). Socioeconomic disadvantage and adolescent women's sexual and reproductive behavior: the case of five developed countries. Family Planning Perspectives 33(6): 251-258.

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological bulletin*, *139*(1), 213–240. https://doi.org/10.1037/a0028931

Taylor, C. 2018. The reliability of free school meal eligibility as a measure of socio-economic disadvantage: evidence from the Millennium Cohort Study in Wales. *British Journal of Educational Studies* 66(1): 29-51.

Weckström, T., Elovainio, M., Pulkki-Råback, L., Suokas, K., Komulainen, K., Mullola, S., Böckerman, P., & Hakulinen, C. (2023). School achievement in adolescence and the risk of mental disorders in early adulthood: a Finnish nationwide register study. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-023-02081-4.

Windle M. (1996). Effect of Parental Drinking on Adolescents. *Alcohol health and research world*, *20*(3), 181–184.

Wu, Y., & Zhang, F. (2023). The Health Cost of Attending Higher-Achievement Schools: Peer Effects on Adolescents' Academic Performance and Mental Health. In D. Guo (Ed.), *The Frontier of Education Reform and Development in China* (pp. 12). Educational Research in China. Springer. https://doi.org/10.1007/978-981-19-6355-1_12.

Zima, B. T., Gay, J. C., Rodean, J., Doupnik, S. K., Rockhill, C., Davidson, A., & Hall, M. (2020). Classification System for International Classification of Diseases, Ninth Revision, Clinical Modification and Tenth Revision Pediatric Mental Health Disorders. *JAMA pediatrics*, *174*(6), 620–622. https://doi.org/10.1001/jamapediatrics.2020.0037

Table 1. Crosstabulations between socio-economic quartile and Key Stage 2 achievement quartile

(a) Any reason

	Low SES	Q2 SES	Q3 SES	High SES
Low achievement	117.7	73.8	64.1	60.3
Q2 achievement	107.2	64.8	56.4	51.1
Q3 achievement	98.5	58.0	51.9	47.0
High achievement	86.0	52.0	46.2	41.6

(b) Alcohol / drugs

	Low SES	Q2 SES	Q3 SES	High SES
Low achievement	46.2	24.7	20.3	18.3
Q2 achievement	42.0	22.6	18.2	16.0
Q3 achievement	38.1	19.4	16.8	15.1
High achievement	30.8	16.0	14.1	12.7

(c) Eating disorders

	Low SES	Q2 SES	Q3 SES	High SES
Low achievement	3.6	2.3	3.0	3.3
Q2 achievement	3.1	2.6	3.2	3.5
Q3 achievement	3.6	3.0	3.4	3.5
High achievement	4.2	3.6	4.4	4.5

(d) Self-harm

	Low SES	Q2 SES	Q3 SES	High SES
Low achievement	53.9	32.9	28.5	26.5
Q2 achievement	53.0	30.1	26.9	24.0
Q3 achievement	50.4	29.7	26.3	22.2
High achievement	46.9	26.7	23.7	20.1

(e) Pregnancy

	Low SES	Q2 SES	Q3 SES	High SES
Low achievement	128.5	64.5	45.2	29.6
Q2 achievement	93.2	44.2	25.5	16.3
Q3 achievement	64.8	25.3	15.3	9.2
High achievement	34.6	13.1	6.3	3.4

Notes: Figures refer to rate per 1,000 children. Red shading indicates where the rate is higher, green shading is where the rate is lower. Any cell size less than 10 has been suppressed to zero.

Table 2. Difference in outcomes between high achieving children from advantaged and disadvantaged backgrounds correcting for Kelley's paradox

(a) Difference in rate per 1,000 under different assumptions or test reliability

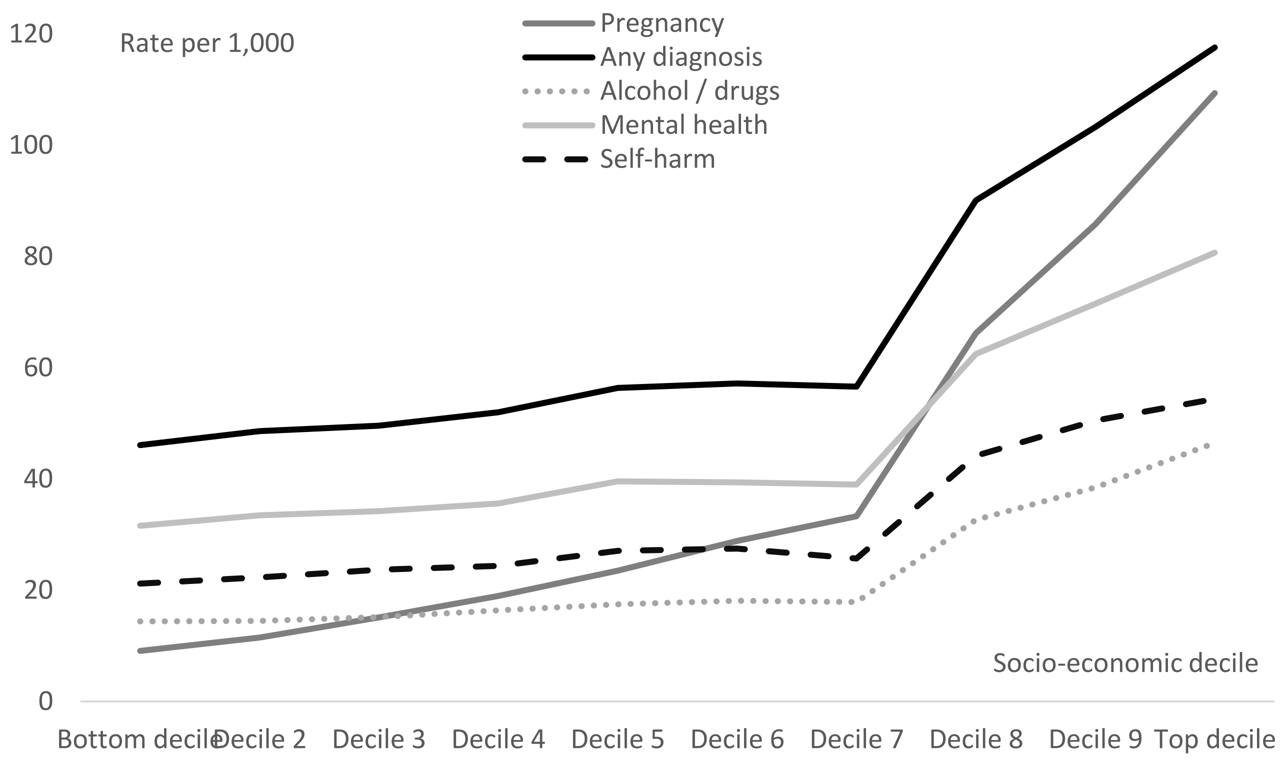
	Rho			SE	N
	1.0	0.7	0.5		
Any diagnosis	42.9	36.5	27.6	1.5	714,937
Pregnancy	28.3	16.7	2.0	1.7	349,709
Self-harm	26.2	23.5	19.5	1.1	714,937
Alcohol / drugs	17.4	14.6	10.7	0.9	714,937
Eating disorder	-0.3	-0.2	-0.1	0.3	714,937

(b) Estimated rate per 1,000 under different assumptions of test reliability

	N	High SES Rate	Low SES rate		
			Rho = 1.0	Rho = 0.7	Rho = 0.5
Any diagnosis	714,937	41.6	84.5	78.0	69.1
Pregnancy	349,709	3.4	31.7	20.1	5.4
Self-harm	714,937	20.1	46.3	43.7	39.7
Alcohol / drugs	714,937	12.7	30.1	27.4	23.5
Eating disorder	714,937	4.5	4.2	4.3	4.5

Notes: Figures in panel (a) refer to the difference in the rate of the outcome per 1,000 children between high-achieving children from advantaged and disadvantaged socio-economic backgrounds. SE refers to the standard errors. Figures in panel (b) illustrate the predicted rate per 1,000 children for those from advantaged and disadvantaged socio-economic backgrounds. Estimates presented under three different assumptions of the reliability of the test used to define the high-achievement group (Rho = 1, 0.7 and 0.5). SE = standard error. N = Number of observations.

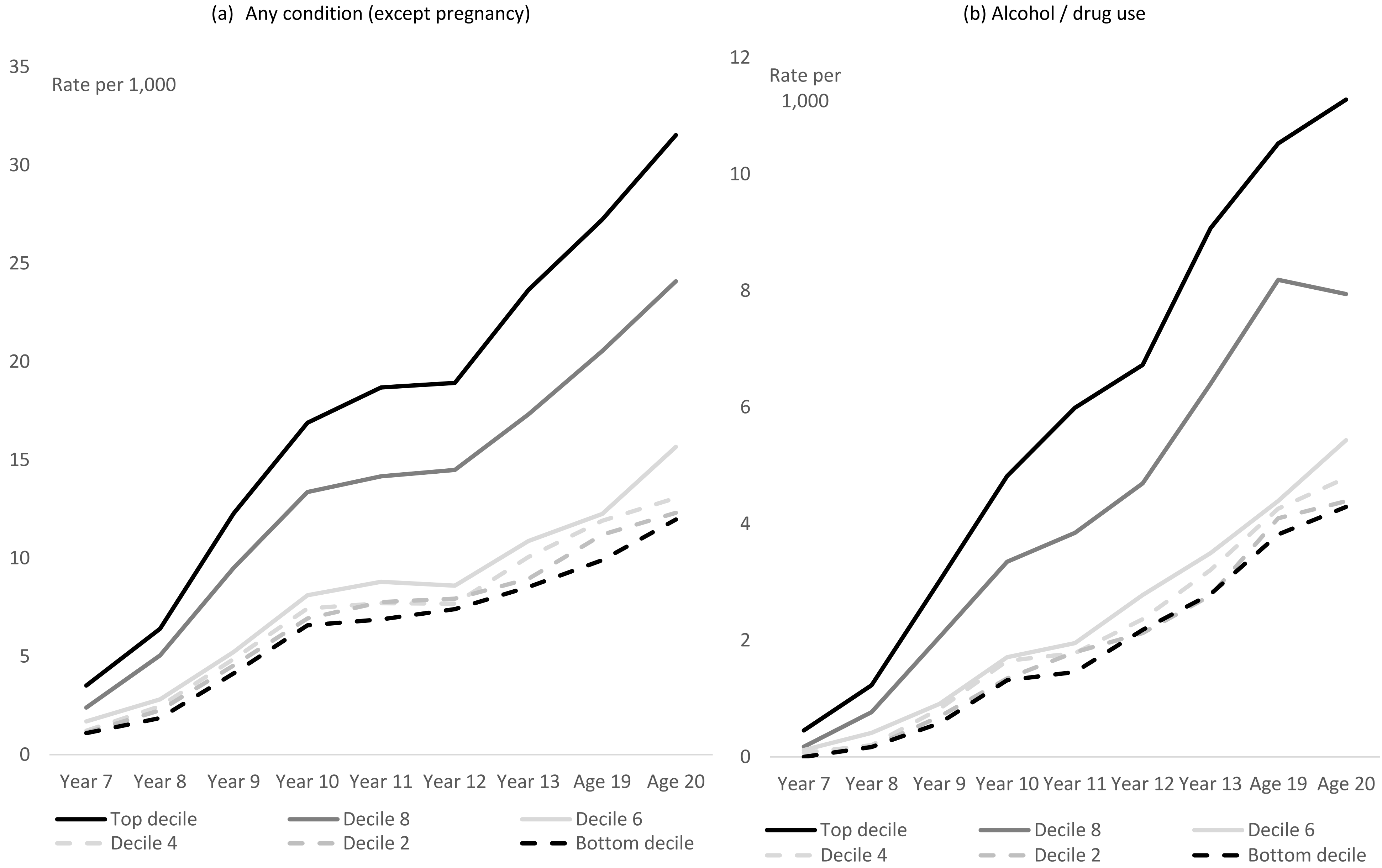
Figure 1. Crosstabulations between socio-economic status decile and pupils health outcomes



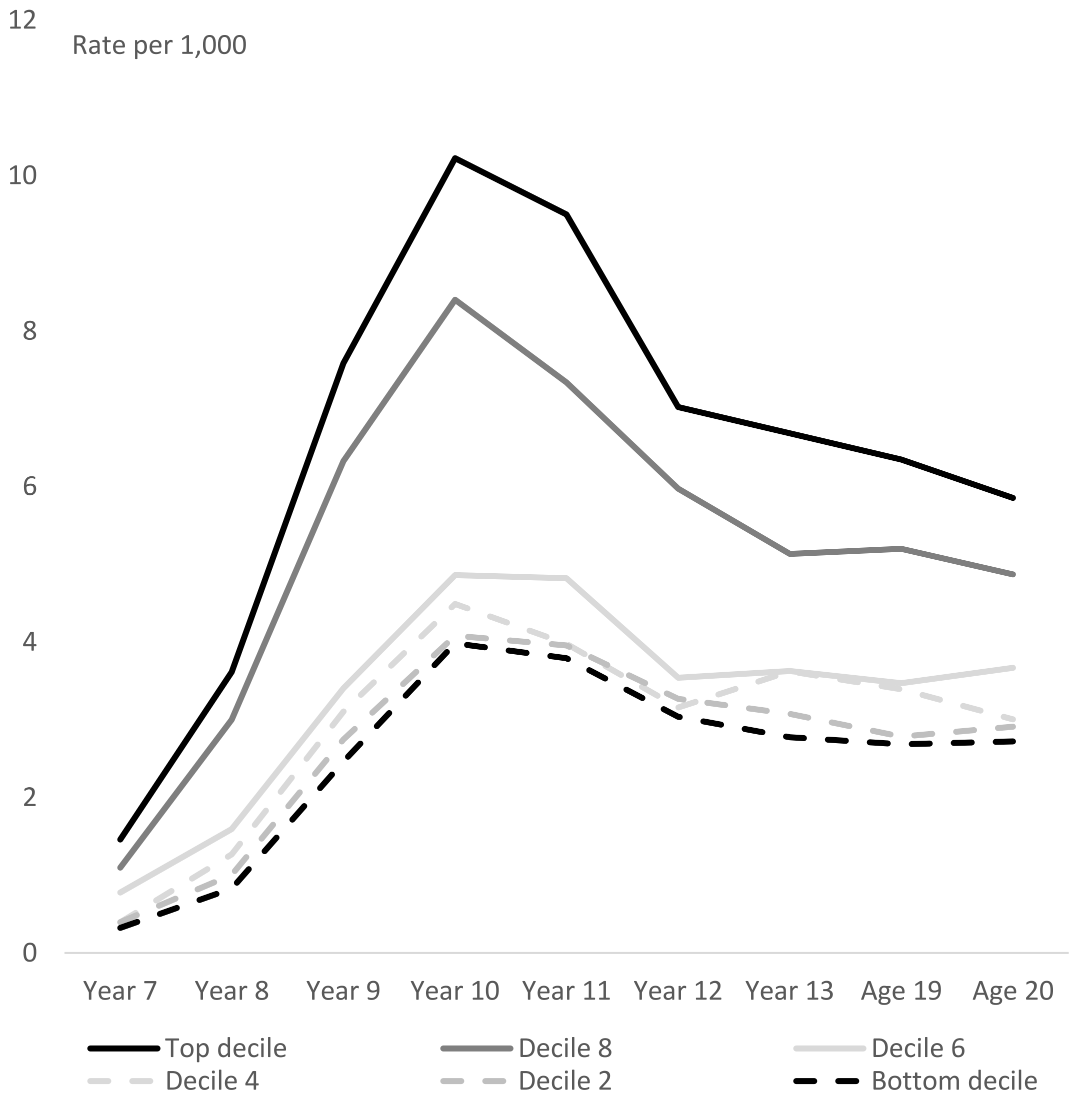
	Pregnancy	Any diagnosis	Alcohol / drugs	Eating disorder	Mental health	Personality disorder	Self-harm	Other
Bottom decile	9.1	46.1	14.4	3.9	31.6	4.1	21.2	2.1
Decile 2	11.5	48.6	14.5	3.9	33.5	3.8	22.3	2.4
Decile 3	15.1	49.6	15.2	3.5	34.2	4.3	23.7	2.5
Decile 4	19.0	52.0	16.4	3.6	35.6	4.5	24.4	2.4
Decile 5	23.5	56.4	17.5	3.4	39.6	4.9	27.1	2.8
Decile 6	28.9	57.2	18.1	2.9	39.4	4.7	27.5	2.9
Decile 7	33.3	56.6	17.9	2.5	39.0	4.7	25.7	2.8
Decile 8	66.2	90.1	32.7	3.5	62.5	8.5	44.2	4.5
Decile 9	85.8	103.3	38.5	3.5	71.5	10.3	50.5	5.4
Top decile	109.4	117.6	46.5	3.5	80.7	12.2	54.4	6.6

Notes: Figures refer to the number of hospital contacts per 1,000 children. Shading should be read vertically within columns. Darker shades of red indicate a higher rate of hospitalisation. Any cell size less than 10 has been suppressed to zero.

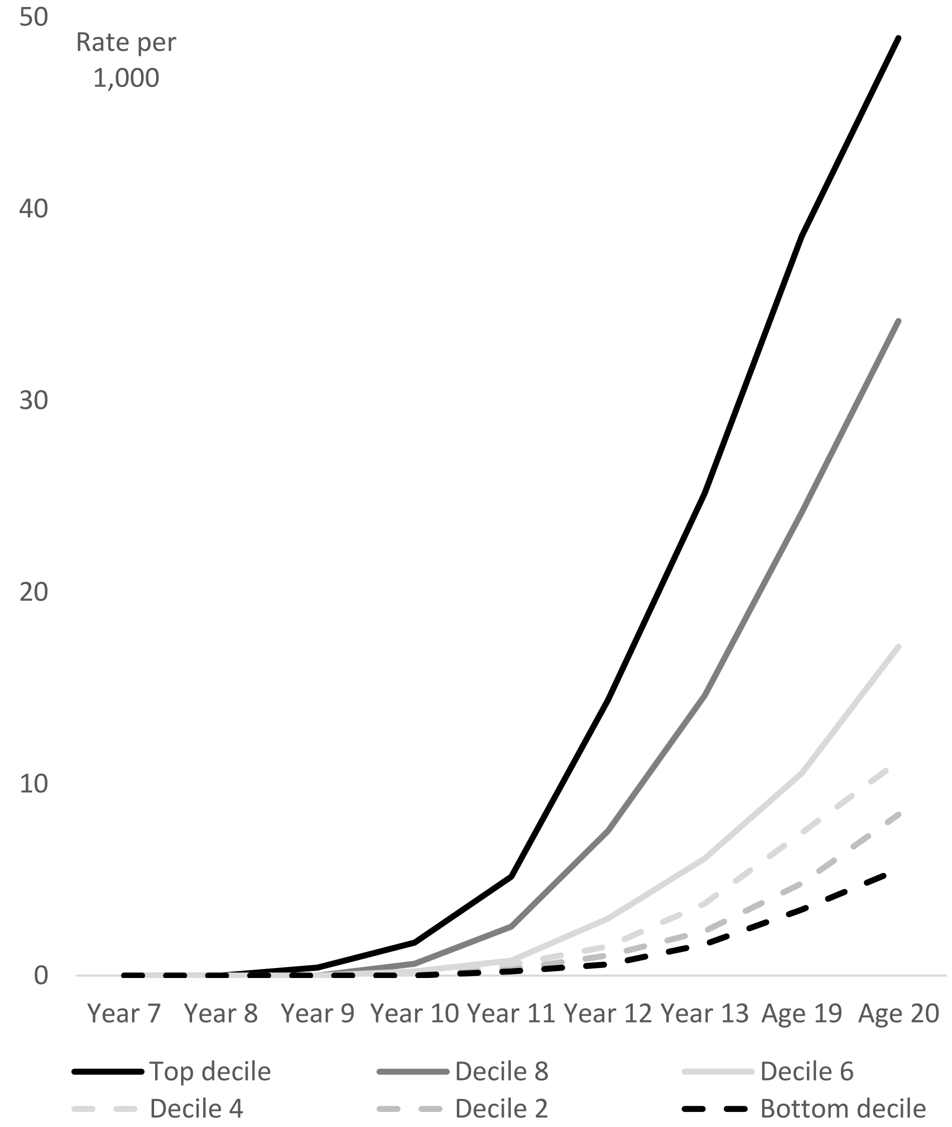
Figure 2. The rate of hospitalisation by socio-economic status decile by age



(c) Self-harm

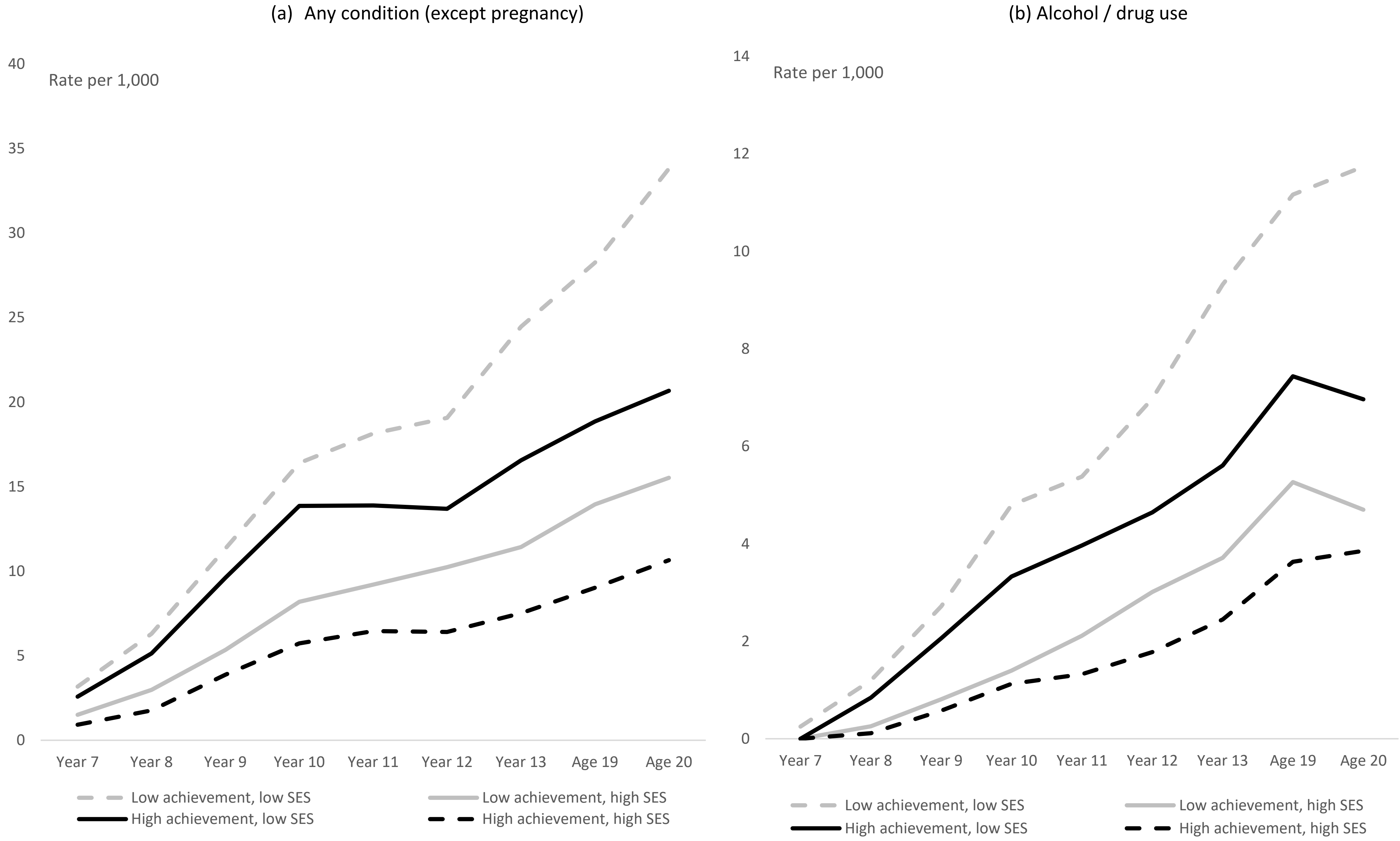


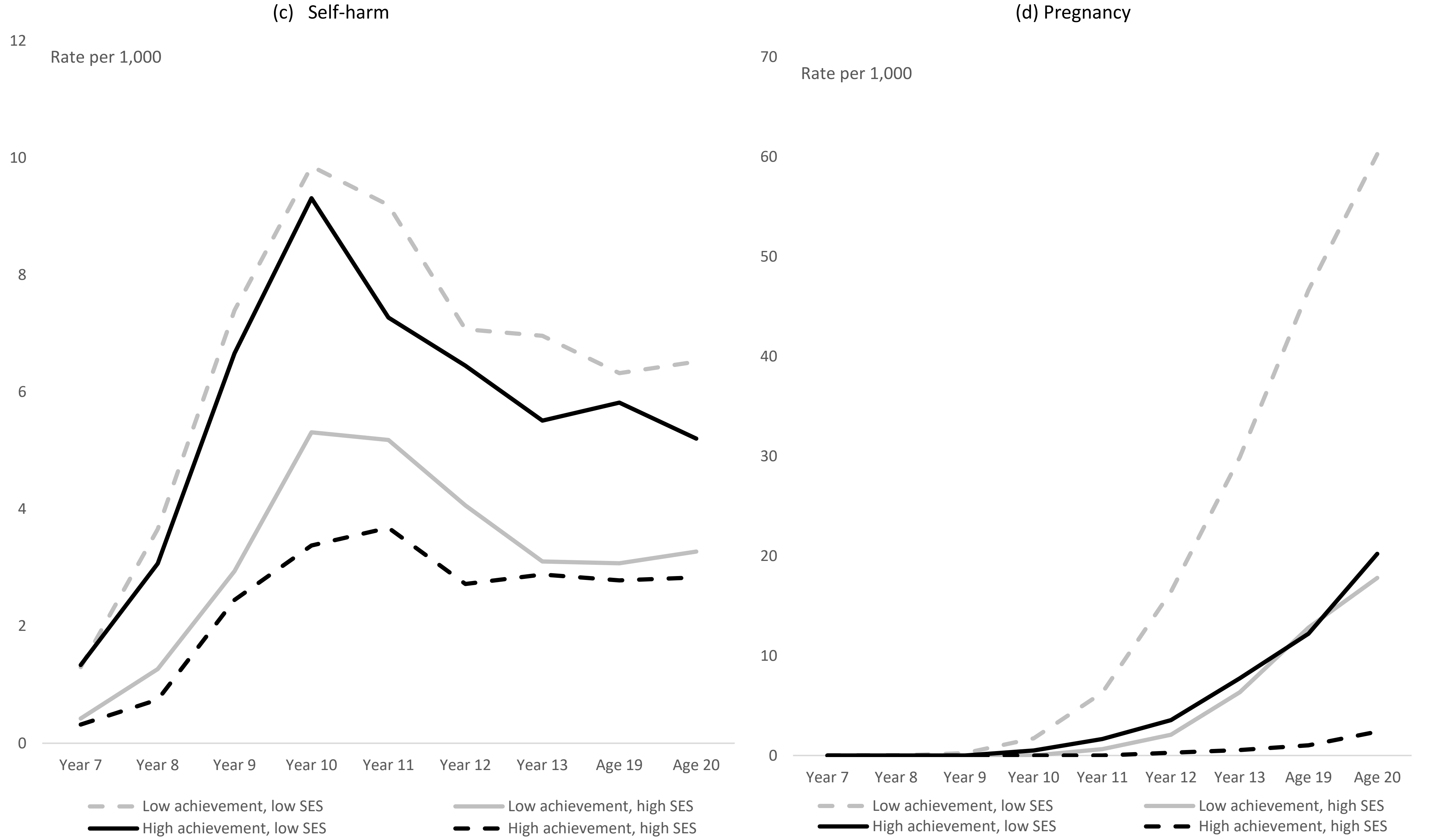
(d) Pregnancy



Notes: Figures refer to the rate of hospitalisation per 1,000 children. Results presented for selected socio-economic status decile. Any cell size less than 10 has been suppressed to zero.

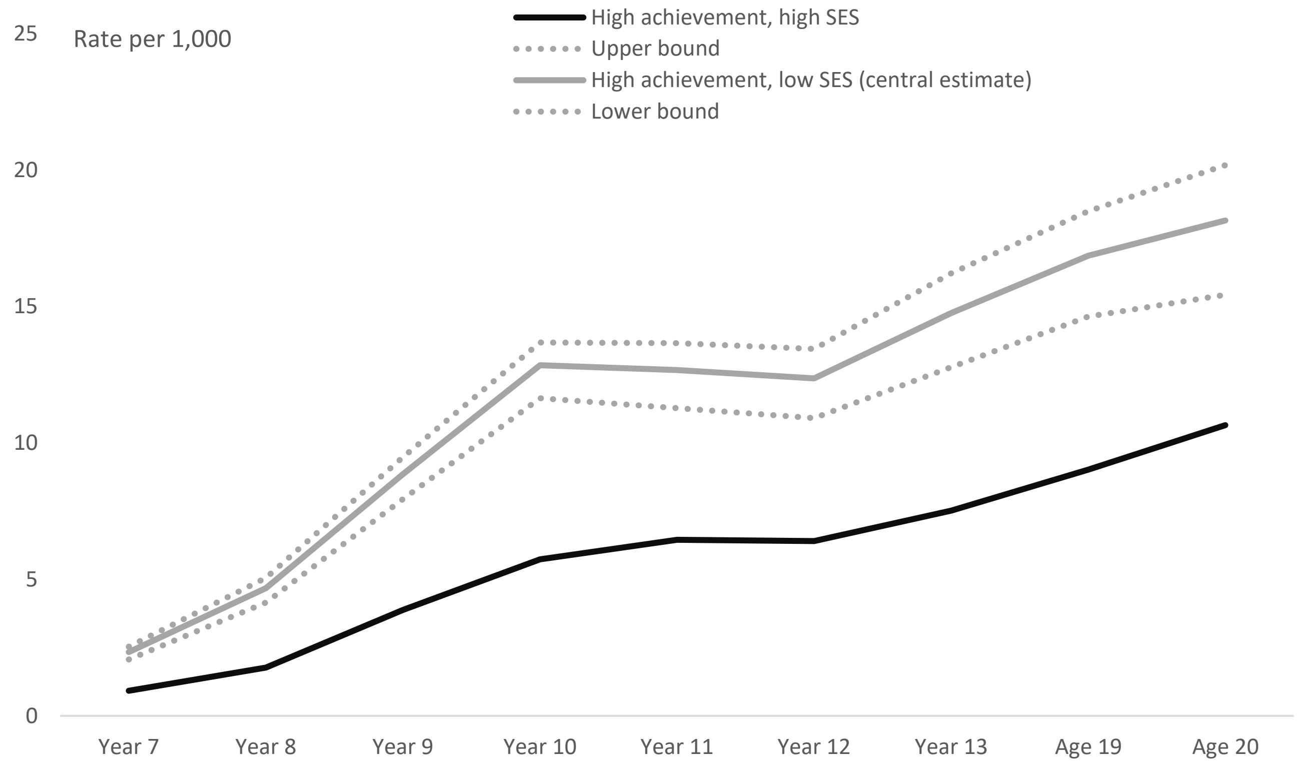
Figure 3. The rate of hospitalisation for children from different socio-economic backgrounds with different levels of achievement





Notes: Low/high achievement refers to the bottom/top quartile of Key Stage 2 test scores. Low/high socio-economic status refers to the bottom/top quartile of the socio-economic status scale. Figures refer to the rate per 1,000 children. Any cell size less than 10 has been suppressed to zero.

Figure 4. Trajectories in hospitalisations due to any risky health behaviour between high-achieving children from different socio-economic backgrounds. Sensitivity of estimates to Kelley's paradox.



Notes: Black line = high-achieving children from advantaged socio-economic. Grey line = Central estimate for high-achieving children from disadvantaged backgrounds, assuming a test reliability of 0.7. Dashed grey lines refer to estimated upper and lower bounds, assuming test reliability of 1.0 and 0.5 respectively. See Appendix H for analogous graphs for alcohol consumption, self-harm and pregnancy. Any cell size less than 10 has been suppressed to zero.