Adolescent School Bullying Victimization and Later Life Outcomes

EMMA GORMAN,†,‡ 🆔 COLM HARMON,‡,§ 🆔 SILVIA MENDOLIA,‡,¶ 🆔
ANITA STANEVA†† 🆔 and IAN WALKER‡,‡‡ 🆔

†University of Westminster, London, UK
‡IZA, Bonn, Germany (e-mail: ian.walker@lancaster.ac.uk)
§University of Edinburgh, Edinburgh, UK
¶University of Wollongong, Wollongong, Australia
††Griffith University, Nathan, Queensland, Australia
‡‡Lancaster University Management School, Lancaster LA1 4YX, UK (e-mail:
ian.walker@lancaster.ac.uk)

Abstract

We analyse the consequences of experiencing bullying victimization in junior high school, using data on a cohort of English adolescents. The data contain self-reports of five types of bullying and their frequency, for three waves, when the pupils were aged 13–16 years. We assess the effects of bullying victimization on short- and long-term outcomes, including educational achievements, income and mental ill-health at age 25 years using a variety of estimation strategies – least squares, matching and inverse probability weighting. The detailed longitudinal data, linked to administrative records, allows us to control for many of the determinants of child outcomes that have been explored in previous literature, and we employ comprehensive sensitivity analyses to assess the potential role of unobserved variables. The pattern of results suggests that there are quantitatively important detrimental effects on victims. We find that both type of bullying and its intensity matter for high-stakes outcomes at 16 years, and for long-term outcomes at 25 such as mental health and income.

I.   Introduction

Being bullied at school is thought to be a widespread phenomenon that harms many children.[1,2] However, there is relatively little quantitative research into the wider and longer term effects of having been bullied as an adolescent, and little research has explored the implications of the heterogeneous nature of bullying.[3] This paper aims to quantify the impacts of bullying victimization on a range of important life outcomes. We analyse several schooling outcomes, including high-stakes examinations at the end of compulsory schooling, at age 16, that track pupils into academic or vocation curricula to age 18 and into university. We also analyse wider outcomes at age 25 years: the effects on income; unemployment; and a mental (ill) health index. In addition to a binary definition of the bullying, we pursue wider definitions: we use factor analysis to create a summary variable capturing the variation in the type and frequency[4] of bullying and we construct a multivalued categorical treatment which allows the effects to differ by type and intensity.

We use statistical methods which rely on a selection on observed variables assumption, paired with a comprehensive range of sensitivity analyses and falsification tests. In particular, we use least squares to adjust for observable factors, as well as matching and weighting methods to reduce the effects of functional form assumptions – employing propensity score matching (PSM) where we consider a single discrete treatment, and inverse probability weighted regression analysis (IPWRA) where we consider multiple treatments. The IPWRA analysis facilitates the estimation of the effects of different types and intensities.

A limitation of our work is that our estimates could be affected by bias due to selection on unobservables. We have data on many of the determinants of bullying identified in the previous literature, and we build a case for depending on a selection-on-observables assumption. Nonetheless, our estimates are not necessarily causal and we recognize that bias from unobserved variables remains a concern. We implement tests which explore the potential role of selection on unobservables, and our application shows that it would take large, likely implausible, levels of selection bias to drive our results to zero. Thus, we feel able to offer this work in the spirit of shining a light on an important but difficult issue, where we currently know very little about the magnitude of the effects. Being able to estimate bounds on the causal effect may show us that bullying is sufficiently important to want to take greater policy action. Moreover, knowing about sources of heterogeneity may tell us how; and what

---

[1]Throughout, we refer to victimization through bullying at school simply as bullying. Bullying in this paper is wholly school based – we do not consider, for example workplace or domestic sources of bullying.

[2]The 2017 edition of the Annual Bullying Survey, a large online non-random 'snowball' survey of young people in secondary schools and colleges across the UK, records 54% of all respondents had been bullied at some point in their lives. According to this survey, one-third of all victims experience social anxiety, one-third experience depression and a quarter of the victims had suicidal thoughts.

[3]A recent comprehensive review of the psychology literature can be found in Ren and Voelkel (2017), and a succinct review of the education literature that focusses on England can be found in Brown (2018). OECD (2017) also provides international evidence.

[4]Olweus (1997) emphasizes the requirement that it is only repetitive actions that count as bullying.

should be the highest priorities. Thus, we report a mosaic of results for a variety of outcomes, reflecting the range of possible definitions of the treatments, estimation methods and control variables. Together, the results suggest that there are important long-run effects of bullying victimization.

II.   Existing bullying literature

Victims of frequent bullying have reported a range of psychological, psychosomatic and behavioural problems – including anxiety and depression, low self-esteem, mental health problems, sleeping difficulties, sadness and frequent pain. Reviews of the work on bullying in the education and psychological literature can be found, for example in Sharp (1995); Ladd *et al.* (2017); Bond et al (2001); Due et al (2005); Arseneault et al (2010); Ford *et al.*, (2017); Woods and Wolke (2004).

There is little economics research on bullying. The seminal work is Brown and Taylor (2008) who use the UK National Child Development Study (NCDS) cohort of children born in a particular week in 1958. The strength of this early contribution to the economics literature on bullying is that it uses a high quality and large cohort study that follows children through school and long into the labour market. The weakness is that it includes many variables that seem likely to be bad controls.

There are few studies that address the issue of causality. Eriksen *et al.* (2014) use Danish data that, like our own analysis, combines administrative registers with a detailed survey, but considers only educational outcomes at age 16 based on tests in language and mathematics skills. Eriksen *et al.* (2014) attempt to estimate the causal effect of elementary school bullying on these age 16 outcomes through an identification strategy based on classroom peer effects. That is, they assume that the proportion of class children whose parents had criminal backgrounds affects other children only through a bullying channel. They report an OLS estimate of earlier bullying of −0.14 of a standard deviation of the grade point average (GPA) at 9th grade. But their IV results, are almost an order of magnitude larger, not smaller as a conventional selection story might suggest. Moreover, it seems likely that having children from extremely challenging backgrounds in the classroom would have an impact on other children in a variety of ways, apart from through a bullying channel.[5] A second paper that considers the endogeneity issue is Sarzosa and Urzua (2020). This uses a South Korea longitudinal survey of 14–18 years olds with matched administrative education data to identify the effects of being bullied at age 15 on mental and physical health, and risky behaviours measured at age of 18. The authors estimate a structural model that assumes that the outcomes and treatment equations are independent from each other once they control for observable characteristics and latent skills. Their bullying definition refers to pupils being severely teased, threatened, collectively harassed, severely beaten, or robbed but is ultimately a single binary variable. Sarzosa and Urzua (2020) show that non-cognitive skills significantly reduce the likelihood of being a victim of bullying and reduce its impact on outcomes. The

---

[5]For example, Carrell and Hoekstra (2010) show that troubled children have a direct negative spillover effect and significantly decrease the reading and math test scores of their peers.

authors estimate that victims have significantly higher incidence of self-reported depression, sickness, mental health issues and stress. For example, being bullied at 15 increases sickness and mental health issues by 0.75 and 0.5 of a standard deviation, respectively, at age 18. More recently, Rees and et. al (2020) assess the effects of bullying victimization on the mental health of 14–18 year olds in the United States of America (US), exploiting state-level variation in anti-bullying laws (ABLs) which aim to reduce bullying in schools. Using a difference-in-difference strategy, the authors find that bullying victimization increased the probability of depression and suicidal behaviours among young women, with the effects being most pronounced among non-white, and lesbian, gay, bisexual or questioning, youth. The effect sizes are large; ABL adoption is associated with a 13–16% reduction in suicides among young women aged 14–18 years. These findings suggest that school bullying is extremely detrimental to the mental health of young people in the US context – especially those in marginalized groups.

III.   Data and specification

We use a large representative cohort study of English children, born in 1989/90, who have been followed from age 13/14 to age 25 years, at which point educational attainment has largely been completed and labour market outcomes are recorded.[6] The data are known as *Next Steps*, and previously as the Longitudinal Study of Young People in England, LSYPE (University College London, 2017a, 2017b).[7] LSYPE covers family, education and labour market variables, and covers sensitive issues, such as risky behaviours and personal relationships. LSYPE selected observations to be representative of the English population, but specific groups were oversampled – in particular, youths from low socioeconomic backgrounds and minorities. Sampling weights are included in the dataset and more details can be found in Centre for Longitudinal Studies (2018) and Anders (2012).

   The survey started in 2004 when the young people were at the age of 13/14 (in school year 9). In the first wave, around 15,000 young people were interviewed across more than 700 schools. The survey followed these individuals for seven years (age 14–21) and then re-interviewed them in Wave 8 at age 25. Non-response in the first wave was approximately 25%, and thereafter there was approximately 10% in each subsequent wave. There was then a four-year break until wave 8 (age 25) – when a lot of new household formation occurs, which contributed to a further drop. There does not seem to have been any substantial attrition as children completed compulsory schooling or when the survey moved to mixed collection methods.[8] The survey data

---

[6]The details of the educational context in England is provided in the Web Appendix.

[7]The Wave 8 survey sought consent from LSYPE participants to allow further administrative data matched to LSYPE. We intend to return to this issue if such a longer-term follow-up of the LSYPE cohort becomes successful. We will also pursue further outcomes that will be included in Wave 9 at age 31 which will become available in 2022.

[8]While it is not possible to completely rule out the possibility of bias due to attrition, because it relates to unobserved selection, we do conduct analyses with and without the survey weights provided. This work does not materially differ from the results presented here. Further details are available on request.

are matched to an administrative register known as the National Pupil Database (NPD), which includes the LSYPE sample of that 1990 birth cohort and the detailed histories of educational attainment of all pupils in the cohort.

Outcomes

We study the impact of bullying on the following outcomes – most of which we think of as being long-term ones, but also include the most important proximate high-stakes educational outcomes:[9]

- Having 5 + GCSE passes, including Maths and English, in junior high exit examinations – which is the most important criterion for a post-16 academic track ('*5 + GCSE*')
- Having at least one Advanced-level qualification (or vocational equivalent) which contributes to university admission ('*Any A-levels*')
- Points sum of the best 3 A-level (or equivalent) points, based on the best three subjects[10] ('*Best 3 A-level points*')
- Receiving a university degree ('*University degree*')
- Natural log of weekly income ('*Ln Income*')
- Not in employment, defined as not being an employee or self-employed, and so includes not in the labour force ('*Not employed*')
- General Health Questionnaire, measuring mental ill-health from 0 to 12, where 0 represents perfect health and 12 represents maximum distress ('*Mental health*').

Bullying data

Our bullying data are unusually comprehensive because it consists of five types, seven frequencies (including none) and were collected over three waves of data corresponding to the lifecycle peak in victimization. This data provide the flexibility to define a large number of possible treatments. The data asks pupils (and the main parent) whether the child was a victim of bullying in the last year. In particular, in each of the first three waves, subjects were asked whether they had experienced any of the following five forms of bullying last year:

- Upset by name-calling, including text or email[11]
- Excluded from a group of friends

---

[9]There are several other long-term outcomes in LSYPE that might be interesting to explore but this would diversify into areas that are less central to the interests of economists. Nonetheless, this would make a worthwhile extension to this work.

[10]Points (which are assigned based on the grades achieved) from the best three A-level or equivalent qualifications are commonly used as the basis of admission by most UK HEIs. See: https://www.ucas.com/sites/default/files/2015-uk-qualifications.pdf.

[11]The data predate the much greater penetration of social media channels. LSYPE2, 10 years after this cohort, records that only 10% of such behaviour is mediated through social media.

- Made to hand over money or possessions
- Threatened with violence
- Experienced actual violence.

In addition to type of bullying, the data contain information on frequency: 'every day'; 'a few times a week'; 'once or twice a week'; 'once every two weeks'; 'once a month'; and 'less often than this'. However, estimating large numbers of treatments defined by interactions between types, frequencies and waves using a dataset with a relatively small sample is unlikely to yield precise estimates. We therefore examine appropriate ways of creating summary measures that seem acceptable to the data. In preliminary OLS estimation, available on request, we use nested testing to aggregate types and intensities to achieve a statistically acceptable specification that would be sufficiently parsimonious to allow estimation using a number of methods. The first definition of a treatment is a binary variable equal to one if a child has experienced any bullying across the three waves, and zero otherwise. The majority of the existing quantitative literature uses just one variable to define bullying, and this treatment provides a baseline specification that is comparable with previous studies. Second, we define a richer summary measure, using factor analysis, which combines information on type and frequency of bullying over the three waves. This imposes no structure on the data but allows us to explore the effects using a variable that captures the complexity of the victimization. The rationale behind these variables is as follows. Rather than imposing constraints on the raw data to generate more parsimonious specifications, we first take a data-driven approach using exploratory factor analysis.[12] We conduct the factor analysis on the frequency of bullying variables, which are distinct by type and wave. We find evidence of just one common factor which we interpret as a measure of cumulative bullying intensity.[13] This score is standardized to have a mean of zero and a standard deviation of one, which allows us to interpret subsequent results in terms of a standard deviation of the bullying intensity. This approach extracts the variation available by type, frequency and wave in a pragmatic, but data-driven, way.

Finally, we define a multivalued categorical variable to capture potential heterogeneity in treatment effects. This variable aims to allow different effects by type and intensity of bullying. We first reduce the number of treatments by collapsing the number of types to two, by combining the three types that relate to violence (actual violence, threatened violence and demanding money or belongings under duress) and collapsing the two non-violent types (name calling and social exclusion) into one. This is largely a practical matter to preserve cell sizes. We justify this aggregation on the grounds that that some types, for example, extortion, have a very low incidence so the data would be unlikely to have the power to detect small effects on outcomes, and the variables in these grouping are naturally correlated: extortion usually occurs

---

[12]Factor analysis is commonly used when using data sets with large numbers of observed variables that are thought to reflect a smaller number of underlying latent variables.

[13]These are found using standard procedures according to which only factors with eigenvalues greater than or equal to one should be retained. See Fiorini and Keane (2014) for a similar application. The first factor explains 73% of the variance. We tried oblique rotation techniques to allow the factors to be correlated but the rotation did not affect the estimates.

TABLE 1

Nine categories of the multivalued treatment

		Non-violent		
		None	*Moderate*	*High*
Violent	None	72%	10%	3%
	Moderate	3%	3%	2%
	High	1%	1%	3%

Notes: Cell percentages do not add to 100% due to rounding.

because of violence, or some implied threat. While the frequency variable is originally reported on a six-point scale from 'every day' to 'less than once per month', we impose cardinal interpretations by assigning the number of school days corresponding to each level, ranging from zero to 200 school days.[14] Therefore, after collapsing into two bullying types, we create two continuous variables by summing the total instances of violent and non-violent bullying instances across the three waves. For example, because for each of the two non-violent types of bullying there are a maximum of 200 instances in each wave, the maximum number of non-violent instances across the three waves would be 1200. This restriction does not focus on heterogeneous effects in the timing of bullying, but rather measures the *cumulative* effect of being bullied. To capture heterogeneity in the pattern of bullying, we create a multivalued treatment variable summarizing the violent and non-violent frequency variables. This variable takes on nine values indicating each combination of: violent, non-violent, no or little bullying, moderate bullying and high bullying. No or little bullying is defined as a frequency of zero days, or the lowest frequency of 2 days. This means this lowest category is 0 to 4 days for non-violent bullying (2 days multiplied by 2 types) and 0–6 days for violent (maximum of 2 days multiplied by 3 types). High bullying is defined as being in the top quartile of the bullying frequency distribution: experiencing 100 days or more of bullying in a school year. Moderate bullying is the remaining group. Thus, we focus on three definitions of bullying – a *binary variable* indicating whether the pupil has been bullied, of any type or frequency, at any point over the three waves of data (and a corresponding variable based on the parent reports); a *continuous variable* constructed via a factor analysis of the frequency of each bullying type in each wave; and a *multivalued discrete treatment* for each combination of violent or non-violent bullying type, and none, moderate or high cumulative frequency of occurrence over three waves. Table 1 summarizes the nature of this variable.

Summary statistics

The most general sample for analysis is restricted to individuals who participated in Wave 8, to yield long term outcomes, and also participated in Wave 1 and have

[14]Assuming 200 school days in a year, we make the following imputations: 'every day' = 200 instances per annum; 'a few times a week' = 100; 'once or twice a week' = 60 instances; 'once every two weeks' = 20 instances; once a month = 10; and 'less often than this' = 2.

complete data on the most basic set of covariates we use (N = 7,569). As we add further covariates and consider outcomes from various sources in our linked administrative data, the sample reduces. Testing for differences in key characteristics across the different estimation samples does not reveal significant differences (available on request). LSYPE contains survey weights, to adjust for the complex survey design (a function of ethnicity, area deprivation and school type, among other factors) and survey drop-out (modelled as a function of observed characteristics in the data). We may wish to use the weights if we suspect they may be correlated with our treatment effects, that is, that the survey design or survey drop-out may bias our results. In the main analyses, we do not use the weights. However, where we can we have also fitted the models with the survey weights, yielding negligible differences in our parameters estimates (available on request), such that we feel confident using the weights would not alter our findings more generally. However, we do adjust the standard errors for clustering by school.

Summary statistics for the outcomes and the control variables are provided in Table 2. These statistics are unweighted and should not be interpreted as population-representative estimates. Some 45% of children are male; 69% self-report white as their ethnicity, 6% of all children report that English is not their first language; the KS2 and KS3 scores are average points scores from the National Pupil Database (NPD), and are recorded at age 10 and 14 respectively; and 16% of children live with just one of their biological parents. Parents were asked if their child was in their first ranked secondary school – which we include because a child might be more likely to be bullied and have lower achievement, irrespective of bullying, if the child has not been able to gain admission to her most favoured school. 82% are placed in their first-choice school.

The Index of Deprivation included in the analysis is the IDACI (income deprivation affecting children index), a subset of the Index of Multiple Deprivation, measuring the proportion of children aged 0 to 15 living in income deprived families, defined including people out of work, and people with low income (Department for Communities and Local Government, 2015). Locus of control captures individual beliefs about whether life events are mostly internally or externally determined (Rotter, 1966). People with an external locus of control believe that they lack control on what happens in life. On the other hand, individuals with internal locus of control generally believe that life events are mostly caused by their own decisions and behaviours. We measure locus of control using responses to six questions and we use factor analysis to create a continuous index of locus of control. LSYPE includes four questions on working attitudes (see the web appendix Table WA5 for details) and we use factor analysis to create an index of work ethic from these. The parental education variables reflect the rapid expansion that had occurred in HE provision in the late 80s and early 90s so that 37% of the children have gained a HE degree compared to 25% for their mothers – the interviewed 'main parent' is the parent most involved in the child's schooling, and is almost exclusively the mother. We have a wide variety of outcomes. The proportion attaining 5 + GCSE passes, 69%, comes from the NPD data and is matched into the LSYPE data. Whether the individual took any A-levels (or equivalent 'level 3' qualifications), 51% in Table 2; and the sum of the points of the best three

Summary statistics for key variables

	Mean	SD	N
Male	0.45	0.50	7,569
Child's ethnic group			
White	0.69	0.46	7,569
Asian	0.17	0.38	7,569
Black	0.06	0.23	7,569
Other ethnic	0.07	0.26	7,569
English second language	0.06	0.24	7,569
Index of area-deprivation	0.22	0.18	7,030
KS2 average points score	27.43	3.92	6,945
KS3 average points score	34.97	6.39	6,960
Highest parental qualification			
Degree or HE	0.25	0.44	7,569
A-level	0.14	0.34	7,569
GCSE	0.26	0.44	7,569
Low/no qualifications	0.30	0.46	7,569
Age of main parent	43	6.0	7,503
Parents separated	0.16	0.36	7,569
At first choice school	0.82	0.39	7,569
Locus of control	0.05	1.00	5,406
Work ethic	0.13	0.96	6,204
5 + GCSE	0.69	0.46	6,698
Points of best 3 A-levels	228.9	99.40	4,018
A-levels	0.51	0.50	7,569
University degree	0.37	0.48	7,569
Income (£ weekly)	303.4	72.5	7,569
Unemployed	0.10	0.31	7,569
Mental (ill) health	2.30	3.12	7,234

subjects taken (excluding General Studies – a very broad subject that is sometimes taken as a fourth A-level subject) are sourced from the NPD data. We use the letter grade to points conversion scale prevalent at the time. Income is recorded for the individual in wave 8 of LSYPE. Unemployed is defined to include all those not in the labour force (i.e. not self-employed or an employee). Mental (ill) health is measured using the General Health Questionnaire (GHQ), where a higher score indicates poorer health.

Table 3 reports means and standard deviations of key variables by bullying status: whether a child has *never* been bullied, has been bullied in only wave of data, or has been bullied across *multiple* waves of data. Boys are slightly more likely to report being bullied than girls. White families are overrepresented among the repeated bullying category compared with other ethnicities. Children in sole parent families are statistically significantly more likely to suffer multiple instances of bullying compared with pupils in two-parent families. There appears to be little difference in the propensity to be bullied by measures of socio-economic status, such as the area-based deprivation index (IDACI), or parental education level. This makes sense because a key determinant of being bullied is being *different* from those around you, rather than

TABLE 3

Differences in key variables by bullying status

	Never bullied (1)	Bullied once (2)	Repeatedly bullied (3)	p-value from t-test of (3) – (2)
Male	0.44	0.47	0.42	0.00
White	0.62	0.70	0.80	0.00
Asian	0.23	0.17	0.09	0.00
Black	0.07	0.06	0.04	0.00
Other ethnic	0.08	0.07	0.07	0.60
ESL	0.07	0.06	0.04	0.00
Index of deprivation	0.23	0.22	0.20	0.00
KS2 average points	27.54	27.27	27.46	0.15
KS3 average points	35.23	34.62	34.95	0.11
Parent Degree/HE	0.24	0.25	0.28	0.02
Parent A-level	0.13	0.13	0.14	0.59
Parent GCSE	0.25	0.26	0.28	0.16
Parent Low/no qual	0.33	0.31	0.26	0.00
Age of main parent	42.54	42.42	42.44	0.89
Parents separated	0.14	0.16	0.18	0.06
At first choice school	0.82	0.81	0.82	0.85
Locus of control	0.12	0.02	−0.01	0.33
Work ethic	0.21	0.08	0.08	0.97
5 + GCSE	0.74	0.67	0.65	0.35
Points of best 3 A-levels	232.31	227.57	225.04	0.53
Any A-levels	0.53	0.48	0.53	0.00
Has a degree	0.39	0.35	0.35	0.95
ln(Income)	5.67	5.67	5.72	0.00
Unemployed	0.09	0.11	0.12	0.51
Mental (ill) health	1.74	2.44	2.92	0.00

the levels of any particular variable (see Ttofi and Farrington (2011). In contrast, there are substantive differences in outcomes by bullying status, especially mental health, unemployment and income.

In Web Appendix Table WA1, we explore whether the bullying pattern a pupil experiences is associated with a higher likelihood of dropping out of the survey by Wave 8, which shows that, in fact, those who are bullied across multiple waves are slightly more likely to remain until the final Wave. This is in line with the patterns of observable bullying selection in Table 3, in that children from White, and more educated households, are more likely to report being bullied, but with no difference by prior attainment – a pattern which does not align with common patterns of survey drop-out, which typically run in the opposite direction.

Figures 1 and 2 give a sense of the distributions of bullying frequency by type of bullying and wave (among those who report both). Figures 1a,b shows the extensive margin of victimization experience by type – that is, the proportion of girls and boys reporting each type of bullying in each wave. Victimization falls across waves for each type, consistent with the existing literature. It is also clear that name-calling and social exclusion are more prevalent for girls and violence more prevalent for boys.
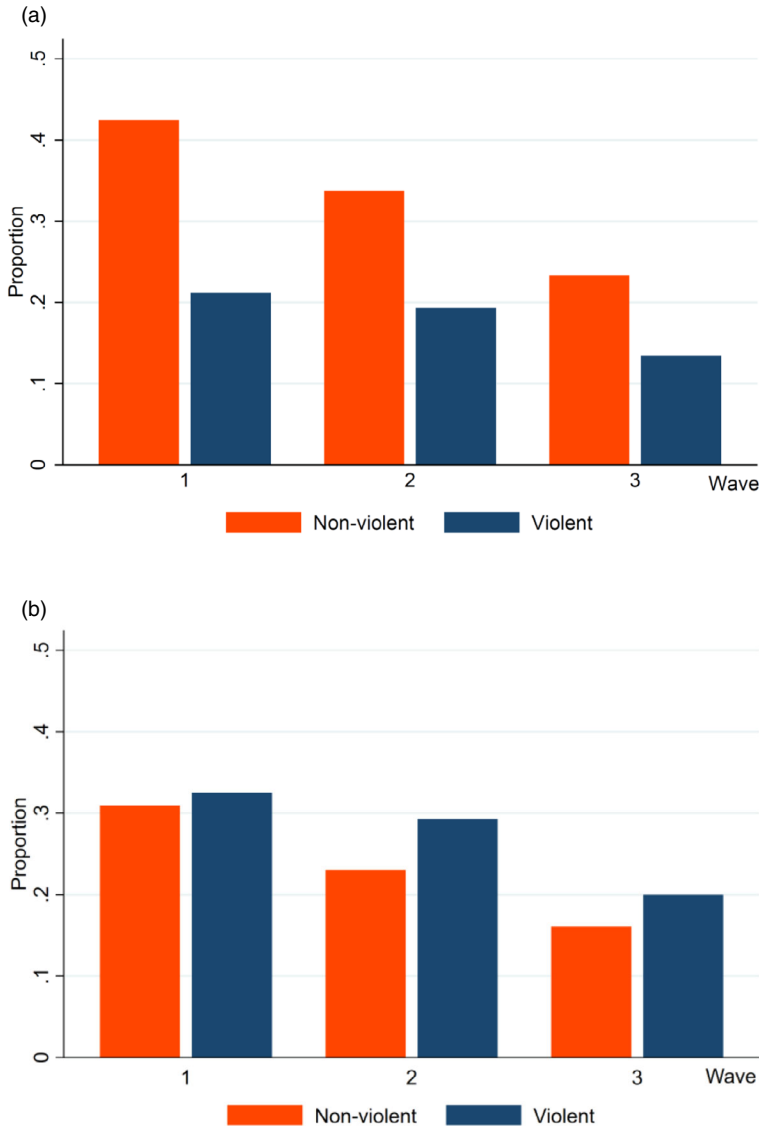
Figure 1. (a) Bullying victimization by wave and type: girls. (b) Bullying victimization by wave and type: boys

Figure 2a,b shows the intensive margin of victimization by type and wave – that is, the average numbers of days the youths report experiencing each type of bullying in each wave. Again, victimization falls over waves and, boys tend to experience more instances, especially of violent types. Exploration of serial correlation in bullying across waves suggested that this was high, for all three main types. For this reason, we feel justified in thinking that frequencies for each type could be aggregated across waves. That is, it may not matter than a bullying instance occurred in Wave 1 or
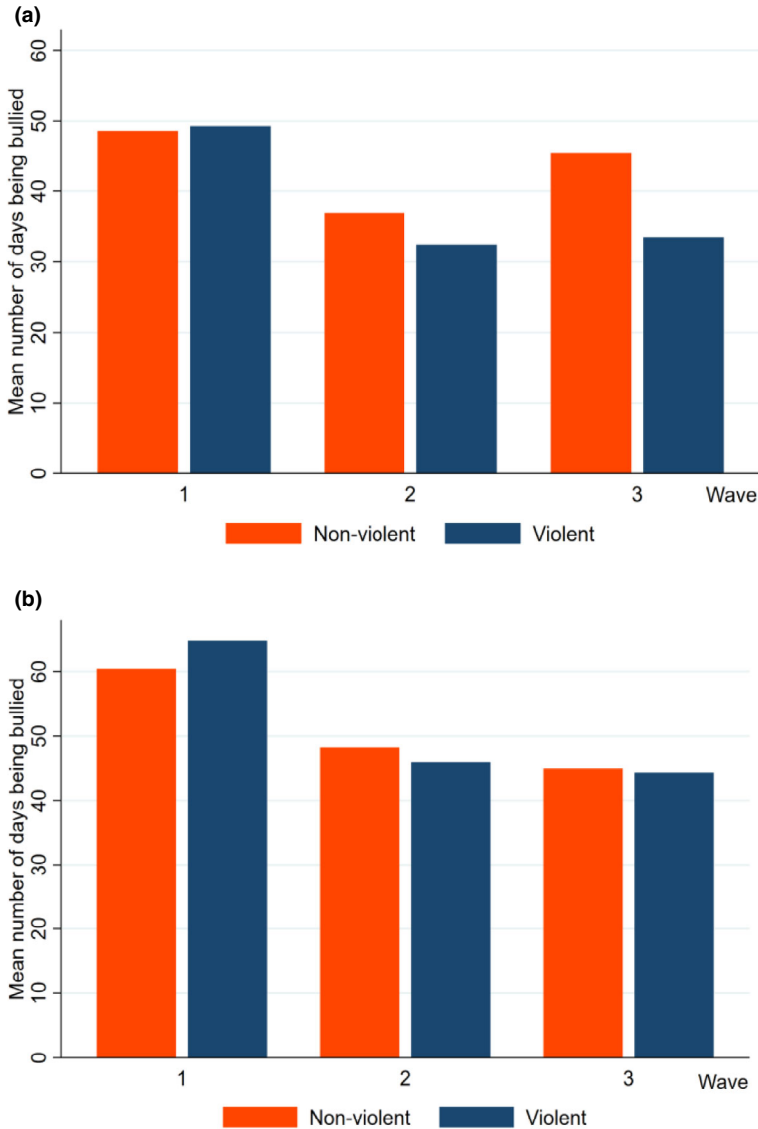
Figure 2. (a) Bullying days by wave and type: girls. (b) Bullying days by wave and type: boys

another wave; rather it appears to be the case that it is only the cumulative bullying experienced that matters, not its distribution by age.

Figure 3 compares the child and parent reports of experiencing bullying. Typically, the child reports show a higher prevalence of bullying. The reports from both child and parents follow a similar downward trend over the three waves reflecting the decrease in bullying as children mature. We show the outcomes associated with each type and frequency of bullying in the web appendix (Figure WA3). We group the days of bullying instances into the three levels defined earlier (none, low, and high), and show, for each intensity group cell, the means for each of our outcomes. These figures
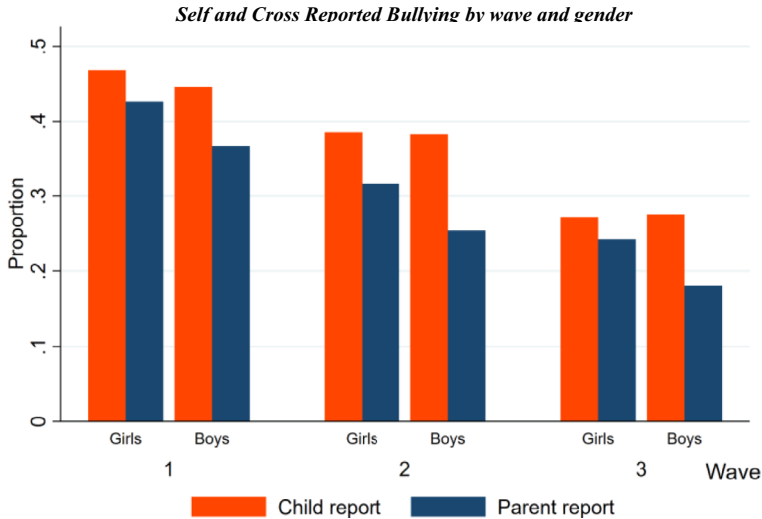
Self and Cross-Reported Bullying by wave and gender

Figure 3. Self- and Cross-Reported Bullying by wave and gender

Notes: Unweighted proportions of cohort members experiencing each type of bullying by survey wave (1,2,3) and gender. 'Non-violent' includes social exclusion and/or name calling, 'Violent' includes threats of violence, actual violence and extortion

show the expected pattern, that increasing bullying intensity is associated with worsening outcomes. This pattern is especially pronounced for unemployment and mental ill-health.

The graphs also foreshadow non-linearities in the effects of bullying: moving from moderate to high bullying is associated with a larger drop in outcomes, compared with moving from no bullying to moderate bullying. This is an issue we return to in our modelling. Finally, the graphs show important differences in the incidence of different types of bullying by gender, but this is not something that we pursue due the reduced estimation efficiency in the gender sub-samples.

IV. Estimation

We explore a range of empirical methods based on various assumptions. We first consider OLS estimates, as a benchmark, then propensity score matching (PSM), and finally multiple treatment effects with inverse-probability-weighted regression (IPWRA).

OLS analysis

We begin by estimating the following simple linear relationship using OLS:

$$Y_{ih} = \mathbf{B}'_{ih}\boldsymbol{\beta} + \mathbf{X}'_{ih}\boldsymbol{\gamma} + \varepsilon_h + \omega_{ih} \tag{1}$$

where Y_{ih} represents one of the several outcomes, observed at age 16, 18 or 25 years depending on the outcome in question, for individual i who attended high school h;

$\mathbf{B}_{ih}$, represents the bullying variable which may be a scalar or a vector, for student i attending high school h;[15] $\mathbf{X}_{ih}$ is a vector of child characteristics (e.g. ethnicity, month of birth, etc.), school characteristics (e.g. school type) and family characteristics (e.g. maternal education and marital status), and ε_h is a school fixed effect while ω_{ih} captures unobservables that vary across i and across h. The inclusion of the school fixed effects allows us to account for unobserved time-invariant school characteristics, which may affect bullying and pupils' outcomes at the same time –for example, the disciplinary regime at the school. Using school fixed effects in many of our models allows us to capture the idea that it is the *relative* characteristics of pupils, compared with one's proximate peers, which are important for determining whether a child is bullied.

In this specification, the coefficients on the $\mathbf{B}_{ih}$ indicators, $\boldsymbol{\beta}$, are the parameters of interest. While the OLS estimator adjusts for observable factors, the resulting estimates do not necessarily warrant a causal interpretation. The plausibility of the conditional independence assumption required for a causal interpretation depends on the relationship between the outcomes and the covariates $\mathbf{X}_i$. As such, it has become common to explore the stability of the parameters of interest by varying the set of control variables $\mathbf{X}_i$. We use two sets of covariates. The first is a parsimonious specification that includes only those variables that seem plausibly exogenous: gender, ethnicity, month of birth, Government Office Region (GOR) and English being a second language (ESL). The second is an intermediate specification which also includes a set of controls which we think of as being predetermined in Wave 1 of the data (age 14): local area deprivation, parental information including age, education, health, income and marital status, test scores at age 10 (KS2) and whether the school was the parent's first choice school. We would be concerned that any further extension beyond this specification would run a risk of including 'bad controls' which would generate biased coefficients.

We implement a number of falsification, or placebo, tests. We assess the effects of the binary bullying variable on variables which should *not* be impacted by bullying if we have adequately controlled for selection into being bullied, that is, they are either determined *before* bullying occurred, or are measured afterward but there is no reason to believe that they should be affected by bullying. Therefore, we expect to not see any significant effects of bullying in this analysis, unless our observed effects of bullying are driven, to some extent, by confounding. Finding appropriate predetermined variables in our data is difficult, but we identify the following candidates: the share of pupils in the school gaining 5 + GCSEs in *2001* (the first wave in the estimation sample is 2004); the Key Stage 2 scores of pupils attending the school in 2001; the deviation of the pupil's average height measured at age 25 from their high school peers; and whether the pupil took either the Math or Science Extension Test at primary school. The rationale for the deviation from average height of peers is to pick up children who may have been relatively small at school, and therefore more likely to be bullied due to their physical attributes. We do not observe pupils' heights while they are at school, only at age 25 years, so we need to make the

---

[15]The dependent variables are only observed once, so although having bullying data for three waves we do not include a t subscript since we accepted the restriction that the corresponding β s did not differ significantly.

strong assumption that relative height has stayed constant. But if we have adequately controlled for the determinants of being bullied, we should not see an 'effect' of bullying on relative height at age 25 years. Finally, taking the Math or Science Extension tests could represent a proxy for being an intellectual or social outlier, as measured prior to high school.

In a second approach to exploring the potential role of unobservable, we use the test proposed by Oster (2019) that indicates the level of selection on unobserved variables, as a proportion of the level of selection on observed variables that would be required to drive the treatment effect to zero – captured by the parameter δ[15]. Results from this test are reported in section VI below, and they provide evidence supporting the credibility of our main estimates.

PSM analysis

We complement least squares estimation with propensity score matching (PSM). Matching offers a number of advantages compared with OLS: increased similarity (balance) in the distribution of covariates between the treated and control groups; explicit consideration of the degree of overlap; and a reduced reliance on a linear functional form. The primary approach we use is kernel propensity score matching. We complement this with a number of alternative estimation methods, to ensure our results are not an artefact of one particular approach: nearest neighbour (NN) propensity score matching, and multivariate distance matching on the Mahalanobis distance (MDM). In Web Appendix Figures WA1 and WA2, we also report a histogram showing the resulting overlap between treated and controls, together with a plot summarizing the balance statistics.

To evaluate the sensitivity of the matching estimates to confounding, we employ the sensitivity analysis developed in (Nannicini, 2007; Ichino et. al., 2008). This sensitivity analysis simulates the effects of a potential binary confounder on the average treatment effect on the treated (ATT). This method is similar in concept to many other sensitivity analyses in the statistics and econometrics literature who also assess the sensitivity to unobserved confounding (e.g. Oster, 2019). One advantage of this specific approach is that is does not require a parametric outcome model, making it suitable to use in a matching context. The idea is that we may suspect that the conditional independence assumption may not hold, given the covariates we observe. However, we might think that conditional on an omitted variable, denoted U, the assumption would plausibly hold. Matching on U, in addition to the vector $\mathbf{X}$, would

---

[15]The assumptions underpinning the calculation of δ can be varied. In particular, the researcher can vary the assumed value of R^2-*max,* the R-squared from a hypothetical regression of the outcome on treatment and both observed and unobserved controls. The default option is to set this as 1, which may not be plausible in situations where it is inconceivable that one might be able to explain all the variation in the outcome. A rule of thumb proposed in Oster (2019) is to set R-*max* equal to 1.3 times the R-squared from a regression of the outcome on the treatment and observed control variables (denoted $\tilde{R}$). The suggested cut-off to define an 'acceptable' level of selection is an estimate of δ (calculated using R-*max* $= 1.3 * \tilde{R}$) that exceeds 1. This was the level that was found to be consistent with that observed in a sample of papers using RCTs in Oster (2019). Therefore, we report δ based on this level of R-*max*.

allow us to obtain a consistent estimate of the ATT. By specifying the joint distribution of U, the binary treatment, denoted B, and the outcome, denoted Y, we can compute the 'unbiased' ATT, which accounts for the confounding effects of U. We can compare this to our original, which does not adjust for U, to assess the difference made by accounting for the unobserved covariate.

To operationalize the method, one needs to specify the distribution of a hypothesized U, in relation to B and Y. Equation (2) highlights the maintained simplifying assumption that U is binary and independent of **X**.

$$p_{bj} \equiv pr(U = 1 | B = b, Y = j) = pr(U = 1 | B = b, Y = j, \mathbf{X}) \qquad (2)$$

After specifying p_{bj}, the relevant value of U is assigned to each observation, depending on which category of b,j they are in, and U is included in the calculation of the ATT as an additional covariate. For a given set of parameters, the matching procedure is performed multiple times with varying draws of U, and the estimate of the ATT is the average over the estimates in each simulation.

The first way we operationalize this is to specify U such that the unbiased effect would be driven to zero, and then assess the substantive plausibility of such a confounder. A second way to operationalize this is to specify U to mimic the distribution of some observed confounder which may represent a more plausible scenario.[16] We choose three such variables to explore this: the 'sole parent family' variable; the 'English second language' variable; and a binary variable. which we call 'outlier', that indicates being in either the top or the bottom decile of the Key Stage 2 distribution in their school (i.e. compared with being in the middle of the distribution as the base category). We choose these variables as it seems plausible that they may possibly affect both the probability of being bullied and the outcomes. We then assess the extent to which these hypothetical confounders would reduce the estimated treatment effect.

To assess the economic plausibility the hypothetical confounder U, when specified to reduce the treatment effect to zero, we report both two types of odds ratios: the *selection effect* and the *outcome effect* (Nannicini, 2007; Ichino et. al., 2008). The *selection* effect quantifies the degree to which the posited unobserved covariate increases selection into being bullied: specifically, the odds of being bullied associated the binary confounder taking the value one, divided by the odds of being bullied associated the binary confounder taking the value zero. The *outcome* effect quantifies the degree to which the posited unobserved covariate increases the average outcome: specifically, the odds of a binary outcome associated with having the confounder taking the value one, divided by the odds of a binary outcome associated with having the confounder taking the value zero. The idea is that if an unobservable must have implausibly large selection and outcome effects to materially change our results then this would provide evidence supporting the robustness of our results. Results from these tests are presented below, as appropriate, and generally confirm that it would take implausible levels of selection on unobservables to invalidate the main findings.

---

[16]We use the user-written Stata program sensatt to implement these procedures.

Treatment effects with IPWRA analysis

The OLS and PSM analysis so far has employed a simple binary treatment. To improve on this, we also consider a continuous treatment constructed using factor analysis on the frequency of each type of bullying in each wave. Beyond this data reduction approach we consider multiple treatments defined by the varying intensities and types of bullying. We examine the role of different types of bullying using inverse probability weighted regression adjustment (IPWRA) treatment effects estimation based on Imbens and Wooldridge (2009) and its implementation in Cattaneo *et al*. (2010).[17] We use IPWRA to explore the effects of a multivalued treatment taking nine values: each combination of no bullying, low bullying and high bullying frequency, for two types of bullying (violent and non-violent).

Specifically, the probability of 'treatment' (in this context, having a certain combination of violent/non-violent and low/high frequency bullying) is estimated using a multinomial logit specification. The inverses of these predicted probabilities are used as weights in a second-stage regression-adjustment (Wooldridge, 2007; Imbens and Wooldridge, 2009; Wooldridge, 2010). The IPWRA estimator has the so-called 'double robustness property' (Wooldridge, 2007, 2010) in that only one of the two equations in the model must be correctly specified to consistently estimate the parameters of interest. Nonetheless, estimation by IPWRA relies on the conditional-independence assumption in order to identify the effect of bullying on long term outcomes. If we have enough information on the observable differences between youths with and without the treatments, we can heavily weight treated observations that have similar observables to untreated individuals and obtain unbiased estimates of the causal relationship between bullying and long-term outcomes (Mendolia and Walker, 2015).

V. Results

We first present headline results for our two simplest cases: where bullying is a discrete variable corresponding to reporting 'any' bullying; and a continuous variable derived from factor analysis. We then report results where we disaggregate to explore the effects of multiple treatments – violent vs non-violent forms and different intensities.

Headline estimates

Table 4 shows the OLS results for the 'Any bullying' measure, which is typically used in the literature.[18] We report results for boys and girls pooled, with a gender control included (the web appendix Table WA3 provides estimates by gender). OLS results are

---

[17]These effects are estimated using the *teffects ipwra* routine in Stata 15 (STATA Corp, 2017).

[18]A not insubstantial group who report being bullied respond to the frequency question, for at least one of the bullying types, with 'it varies' (n = 885/7,569). While this group are counted as being bullied (bullying = 1) in the discrete definition used to generate the Table 4 OLS results and the Table 7 PSM results, for the analyses which utilize the frequency data (the factor analyses and the IPWRA analyses), the 'it varies' frequency report is set to missing. We take a conservative approach of using complete-case analysis, rather than endeavouring the impute the potential values of the 'it varies' frequency.

reported in Table 4 for effects on having 5 + GCSE passes, taking A-levels, and A-level score (which is a primary determinant of university admission); the intermediate outcome of having a university degree by age 25; and long run outcomes at age 25 (log income, being unemployed and the mental health score). Model 1 includes as covariates: gender, ethnicity, birth month, region (GOR) and English being a second language, along with school fixed effects. Adjusting for these basic controls, we find large detrimental effects of experiencing bullying. The probability of gaining 5 + GCSE passes is reduced by 6.3% points (10% from a mean of 0.69). The probability of staying at school to take a A-levels is reduced by 4.6% points (9.0% reduction from a mean of 0.51), and the A-level points from those qualifications are reduced by about 5 points (5% of a standard deviation). Turning to longer run outcomes, income at age 25 years is reduced by 2.3% (£7 per week of the sample mean). The probability of being unemployed increases by 3.5% points (35% from a mean of 0.10). Perhaps most strikingly, the GHQ mental ill-health index increases by 0.97, a large effect size of about one third of a SD. Evidently, being subject to any bullying, within schools, controlling for basic covariates, is strongly associated with worsened outcomes.

However, these effects may be driven, to some extent, by confounding. Model 2 aims to address this by adding a rich set of relevant controls that are associated with both being bullied and child outcomes. For the GCSE outcome, A-level participation, income at age 25 years, and university degree, adding these relevant controls reduces the effect sizes by about half, and they remain statistically significant (aside from university degree. For example, the coefficient on gaining 5 + GCSEs is now 40% lower (the coefficient is −0.035 in Model 2); the coefficient on staying on in school to take any A-levels is −0.025 points (reduced by 45% compared to Model 1); the points gained from those qualifications is reduced by about −6 points (6% of a SD) in Model 2; the probability of having a university degree is now reduced to a 1% point fall (although not significantly so); and the effect on income at 25 is now reduced by 1% (a £3 per week reduction from the sample mean); and, finally, the GHQ mental ill-health index larger at 0.91, a robustly large effect size of 29% of an SD. However, there may still remain some selection on unobservables, which we explore by reporting the δ parameter proposed in Oster (2019). The estimates of the δ parameter are consistent with an 'acceptable' level of selection according to the rule-of-thumb suggested in Oster (2019). The only exception (for Model 2 results) is log income, which has a negative δ associated with it.[19] Table 4 also presents the bounds on β that are derived from imposing the reasonable (but conservative) assumption that selection on unobservables is equally as important as selection on observables, and sign of the selection effect runs in the same direction for both observables and unobservables (δ = 1).

Table 5 shows the same analysis for the case where bullying is recorded as a continuous variable, based on a factor analysis exercise. There is a very similar pattern of results to Table 4, although the coefficients are not comparable because Table 5 is based on a continuous measure of bullying while Table 4 is a simple dummy variable. Model 2 generally has smaller coefficients than model 1 in both tables, as we might expect. The

---

[19]A negative value of delta is generated when the correlations between observables and the treatment, and between unobservables and treatment, have opposing signs.

TABLE 4

OLS estimates of the effects of 'any bullying' with Oster diagnostics

Dep Var:	5 + GCSE	Any A-levels	Best 3 A-level points	Ln (income)	University degree	Not employed	Mental health
Model 1							
β	−0.063***	−0.046	−4.927	−0.023***	−0.023*	0.035***	0.969***
se	(0.012)	(0.011)	(3.412)	(0.004)	(0.012)	(0.007)	(0.075)
δ	39.07	−5.36	783.10	−4.46	10.12	104.20	66.60
β(δ = 1)	−0.063	−0.054	−4.939	−0.028	−0.021	0.035	0.988
N	6,698	7,569	4,018	7,569	7,569	7,569	7,234
Model 2							
β	−0.035***	−0.025**	−5.880*	−0.010***	−0.011	0.028***	0.911***
se	(0.010)	(0.012)	(3.509)	(0.003)	(0.013)	(0.007)	(0.082)
δ	3.47	2.75	673.70	−36.59	1.68	10.25	11.99
β(δ = 1)	−0.025	−0.016	−5.899	−0.010	−0.004	0.026	0.887
N	6,133	6,413	3,671	6,413	6,413	6,413	6,162

Notes: Robust standard errors, clustered by school, in parentheses. ***$P < 0.01$, **$P < 0.05$, *$P < 0$. School fixed effects are included in all specifications. β = coefficient on bullying; se(β) robust standard error of β; δ (calculated using $R\text{-}max = 1.3 * \tilde{R}$, as explained in STATA's `psacalc` routine) indicates how much selection on unobservables would be required to drive the β to zero, measured proportional to selection on observed variables. Results by gender are reported in Web Appendix Table WA4. We lose precision and generally the boy/girl differences are not significant with the exception of the effect on university degree which is found to be strongly negative for boys and zero for girls.

Oster's δ, as to be expected, generally falls as we move from model 1 to model 2. Yet, in most cases, the results in both tables suggest that, even in model 2, the estimates are unlikely to be zero under reasonable assumptions about possible confounders.

The Oster bounds in Table 5 reflect this conclusion, as the intervals bounded by the treatment effect (assuming no selection on unobservables) and the bias-adjusted treatment effect (assuming equal, positive, proportional selection on observables and unobservables) do not contain zero.

Robustness checks

Table 6 presents results from several falsification tests and show OLS estimates of the association of being bullied with four outcomes: historical information on school performance, and two individual level outcomes – absolute deviation from average peer height and whether the individual took KS2 Maths and Science extension tests[20]. Observing an effect on these outcomes would suggest that we are conflating the bullying effects on long run outcomes with omitted variable bias. Conditioning on the variables listed in Model 2, we do not observe any significant effects on these outcomes. This provides further support for the credibility of our results and for thinking that Model 2 successfully controls for the key determinants of bullying outcomes.

---

[20]The extension paper was introduced in 2001/2 to measure and stretch the most mathematically able 11 year olds (scoring over 90% in the main test). It was subsequently dropped, but was later reintroduced.

TABLE 5

OLS estimates of the effects of 'bullying factor' with Oster diagnostics

Dep Var:	5 + GCSE	Any A-levels	Best 3 A-level points	Ln (income)	University degree	Not employed	Mental health
Model 1							
β	−0.049***	−0.046***	−1.780	−0.018***	−0.027***	0.020***	0.296***
se	(0.008)	(0.007)	(2.930)	(0.003)	(0.006)	(0.006)	(0.049)
δ	26.75	22.60	4.83	−67.67	26.92	−24.51	−18.46
β(δ = 1)	−0.048	−0.045	−1.420	−0.019	−0.027	0.020	0.303
N	4,861	4,890	3,053	4,890	4,890	4,890	4,704
Model 2							
β	−0.014**	−0.023***	−1.203	−0.007**	−0.012*	0.011*	0.297***
se	(0.007)	(0.008)	(2.947)	(0.003)	(0.007)	(0.006)	(0.053)
δ	1.14	2.51	1.14	1.86	1.72	6.51	−49.78
β(δ = 1)	−0.002	−0.014	−0.152	−0.003	−0.005	0.010	0.305
N	4,450	4,464	2,780	4,464	4,464	4,464	4,307

Notes: See Table 4. Here we also find a loss in precision when we distinguish between boys and girls, and we find a strong negative effect on 'Any A-levels' for girls but none for boys.

Apart from the issues associated with identification, the estimates could also be driven by the functional form imposed in the OLS estimation. Therefore, we also investigate propensity score matching. In Table 7, the propensity score findings show a similar pattern to those in Model 1 of Table 4, suggesting that the Table 4 results are not driven by the functional form of the OLS model.

The first row of estimates in Table 8 is copied from the first column of Table 7's PSM estimates. We first consider, in the second row, the plausibility of a binary confounder that would drive our treatment effects to zero. Taking the first outcome in Table 8 as an example, gaining 5 + GCSE passes, we see there would need to be large outcome and selection effects to make this effect completely disappear. The binary confounder U would need to *increase* the odds of being bullied by a factor of at least 4.3 and *decrease* the odds of gaining 5 + GCSE passes by a factor of 0.2. While this may be plausible, looking across the outcomes it seems that the longer run effects are most robust to selection (and so would require the most extreme confounding to reduce the treatment effect to zero). For instance, for mental health, the binary confounder U would need to *increase* the odds of being bullied by a factor of at least 16.3 and increase the odds of being in the top quartile of the mental ill-health distribution by a factor of 23.6. This type of extreme confounder seems an unlikely scenario. To assess more realistic potential confounders, we evaluate the effects of simulated variables that mimic the distribution of relevant observed variables in our data, in relation to the treatment and outcome. The next three rows in Table 8 assess the effects of adding each of our selected simulated variables to be potential confounders in our data: being in a sole parent family; having English as a second language; and being in the top decile or bottom decile of the Key Stage 2 distribution on the child's school. These variables were chosen as being likely to reflect perceived or actual differences from one's classmates, which would both shape the propensity of being bullied and be likely to have direct effects on the outcome.

TABLE 6

OLS linear estimates of 'any bullying' (a binary treatment variable) on predetermined variables (falsification tests)

Outcome:	% pupils 5 + GCSEs in 2001	Average KS 2 points in 2001	Absolute deviation from mean height at 25	Took KS2 Math / Science extension tests
β	−0.179	0.989	0.0013	−0.0034
se(β)	(0.459)	(1.140)	(0.0017)	(0.0062)
N	6,260	5,702	6,478	6,731

Notes: Standard errors are in parentheses. The control variables used in the regressions are from Model 2. For the school-constant outcomes (e.g., % white pupils), school fixed effects are omitted from the specification.

TABLE 7

Propensity score matching estimates of the effects of 'Any bullying'

	ATT	Std. error	N (control)	N (treated)	Total N
5 + GCSE	−0.071[***]	0.010	2,401	3,732	6,133
Any A-levels?	−0.053[***]	0.011	2,526	3,887	6,413
Best 3 A-level points	−7.460[**]	2.757	1,549	2,122	3,671
ln(Income)	−0.017[***]	0.005	2,526	3,887	6,413
Has a degree	−0.022[**]	0.012	2,526	3,887	6,413
Unemployed	0.035[***]	0.007	2,526	3,887	6,413
Mental health	0.960[***]	0.080	2,417	3,745	6,162

Notes: Kernel matching estimation is implemented using `attk` in Stata; se, standard error (bootstrapped with 100 replications). ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$. The covariates included in the propensity score model are from Model 2. The PSM common support graph is available in Web Appendix Figure WA1 and the propensity score balance graphs are available in Web Appendix Figure WA2. Multivariate distance and Nearest Neighbour matching results are available in Web Appendix Tables WA2 and WA3.

TABLE 8

Sensitivity analysis for PSM estimates of the effects of 'Any bullying'

Outcomes:	5 + GCSEs	Any A levels	Best 3 A level points	Ln Income	Not employed	Mental health
ATT assuming unconfoundedness	−0.071	−0.053	−7.470	−0.017	0.035	0.96
U chosen so ATT ≈ 0						
Selection effect	4.258	3.044	3.541	1.834	3.069	16.354
Outcome effect	0.244	0.343	0.335	0.103	11.605	23.599
ATT, *U* mimicking 'Sole parent family'	−0.069	−0.053	−7.450	−0.016	0.035	0.957
ATT, *U* mimicking 'English second language'	−0.071	−0.053	−7.521	−0.018	0.036	0.953
ATT, *U* mimicking 'Outlier in KS2 dist[n]'	−0.070	−0.053	−7.546	−0.017	0.035	0.959
N	6,133	6,413	3,671	6,413	6,413	6,162

Notes: This sensitivity analysis is implemented using the user-written program `sensatt` in Stata. The covariates included in the propensity score model are those from Model 2.

Beginning with effects of the simulated unobserved confounder mimicking the distribution of the sole parent family variable, this would reduce the effect on the GCSE variable by only about 3% (i.e. comparing row 1 with the equivalent row in column 1 of Table 8, −0.071–0.069). Adding the simulated variable to the A-level outcomes determinants the bullying estimates is reduced by only 4% and in the ln (income) equation by about 6%, but it has negligible effects on the other treatment effects. The simulated unobserved confounder mimicking the distribution of ESL again has little impact on the estimated treatment effects, aside from reducing the effect of on mental health by 1%.

Finally, we also examine the simulated unobserved confounder that mimics the distribution of being in the tails of the prior ability distribution. This would reduce the effect on the GCSE variable by about 1.4% and has negligible effects on the other treatment effects. Our conclusion from this analysis is that, overall, scenarios emulating realistic levels of confounding could reduce our treatment effects by between 0% and 6%, depending on which outcome is considered. Therefore, it seems unlikely that our results could be entirely, driven by selection. The type of confounding required for this to be likely to happen appears to be substantively implausible.

Multiple treatments

Finally, we explore the role of type and frequency together using treatment IPWRA estimation of multiple treatments. The aim is to show the merit of viewing bullying as a multivalued treatment problem. Figure 4 summarizes the estimates (available in Web Appendix Table WA5) for the four long run outcomes: (a) university degree, (b) income, (c) unemployed and (d) mental health. The dots are point estimates, while the vertical lines represent 95% confidence intervals. Our results, for each outcome variable, are grouped into three groups, where we look at the effect of increasing violence for a given level of non-violent bullying (grouped by the dashed lines). Thus, in Figure 4a, we see that with, no violent bullying, increasing the intensity of violent bullying decreases the probably of having a degree. Similarly, in Figure 4d, we find that there is (almost) a monotonically increasing adverse effect of violent bullying at any level of non-violent bullying; **and**, looking across groups, we see that as the **non-violent** bullying level rises this (almost) monotonic pattern of increasing effect of violent bullying across the groups of non-violent levels increases successively.

While most of these individual interacted treatment effects are not individually statistically significant from zero, the pattern of these results suggests that the interaction of more bullying of one type conditional on the level of another type is generally to have an adverse effect on outcomes. The estimates for the short-term outcomes are presented in Figure 5. We again visualize the estimates as interactions of more serious bullying and divide these into three groups of successively higher levels of severity. Even with this minimal extension that considers just two types of bullying at three levels of intensity (none, low and high), we find systematic effects of both type and frequency using IPWRA. Especially for the longer run outcomes, it appears that much of the effects is driven through the most intense forms of bullying – high intensity, and violent. Other types and frequencies also have effects, especially for
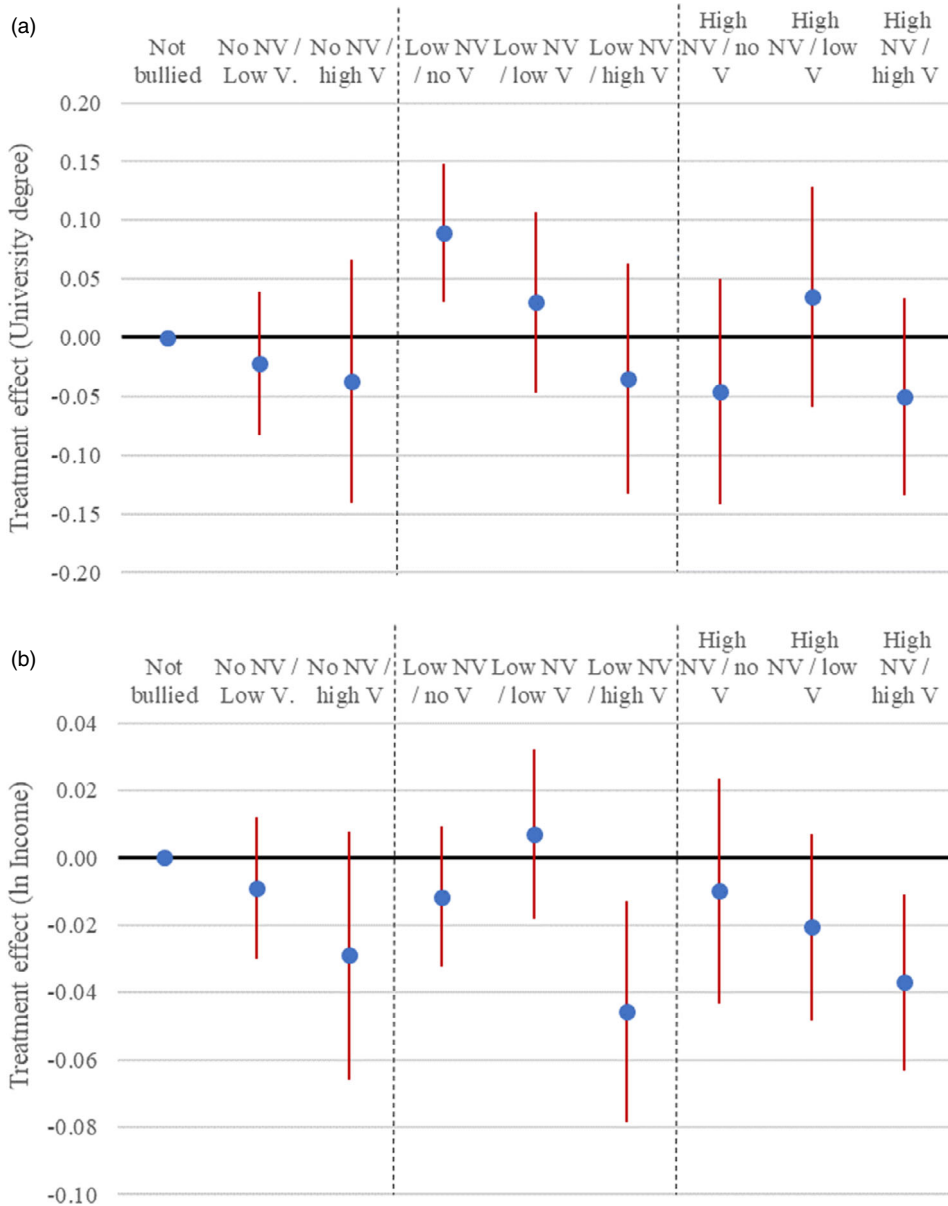
Figure 4. (a) Estimated long-term effects from IPWRA model: university attendance. (b) Estimated long-term effects from IPWRA model: income. (c) Estimated long-term effects from IPWRA model: unemployed. (d) Estimated long-term effects from IPWRA model: mental ill-health
Notes: V = violent, NV = non-violent. Parameter estimates are listed in Web Appendix Table WA5.

mental health where any combination of violent (V) and non-violent (NV) bullying, whether at high or low intensity, have statistically significant adverse effects – raising the mental ill-health count by between 0.5 and 1.5 where the mean is 2.3. The effects on income are large and negative (−4%) only for the relatively small proportion of the
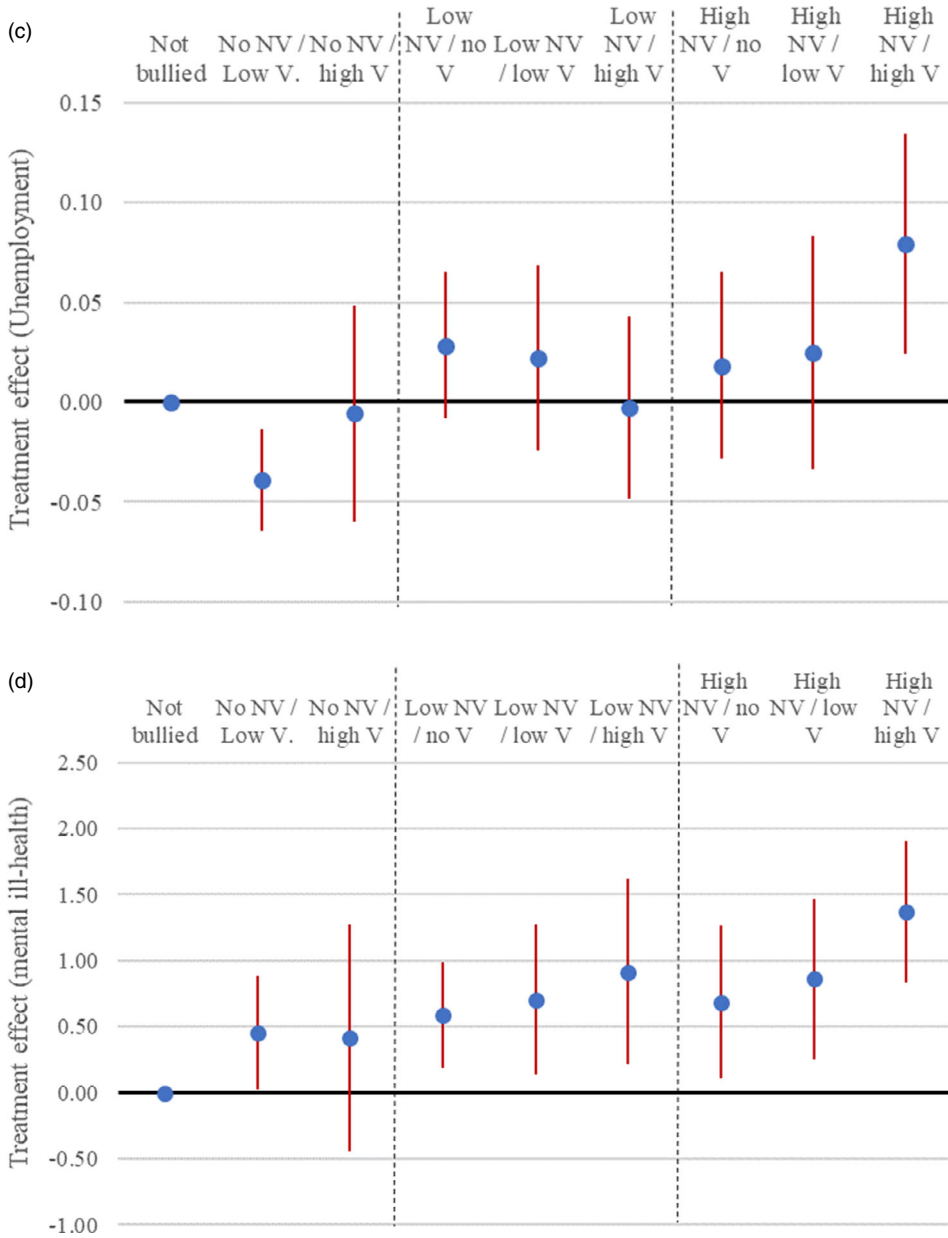
(c)

(d)

Figure 4. (Continued)

population who experience high intensity bullying and either high or low NV bullying. The results in Figure 5 suggest that these effects on income in Figure 4 may stem, in part, from negative impacts of bullying combinations on the probability of attaining 5 + GCSEs or any A level. These results strongly reject the idea that a single treatment is sufficient to capture the complex effects of bullying.
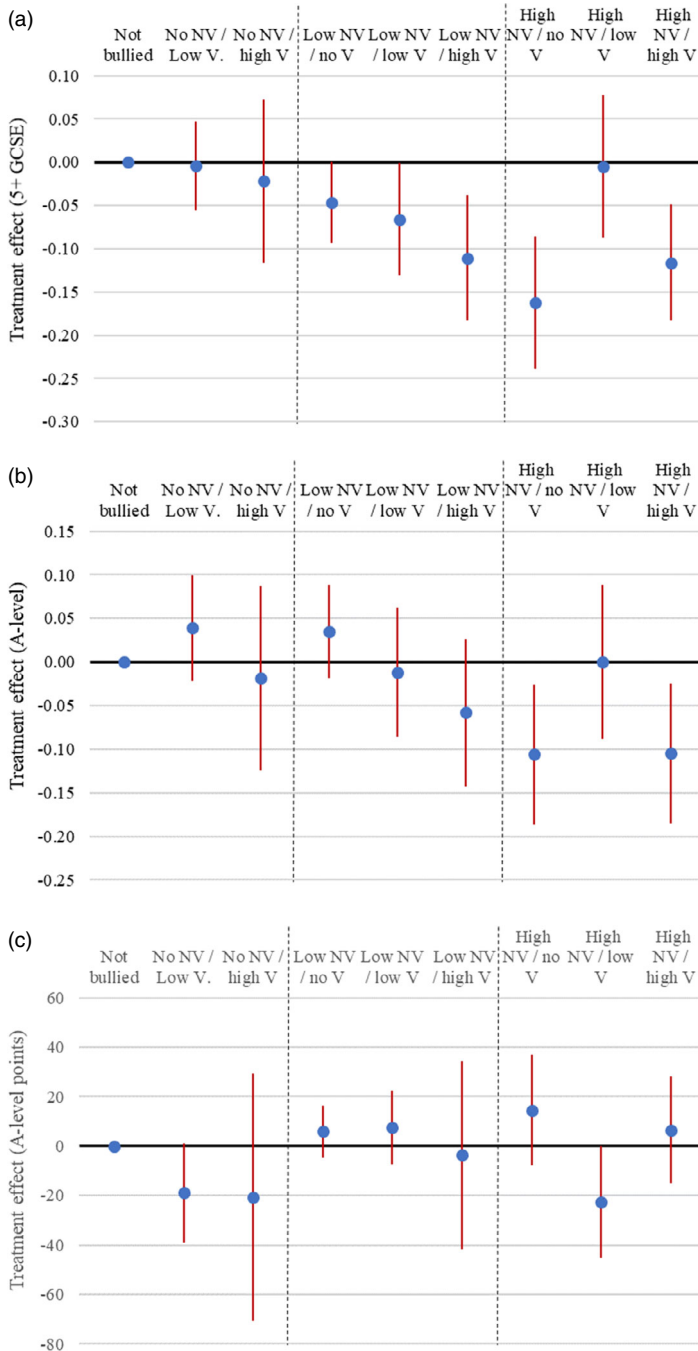
Figure 5. (a) Estimated short-term effects from IPWRA model: 5 + GCSE. (b) Estimated short-term effects from IPWRA model: any A-levels. (c) Estimated short-term effects from IPWRA model: best 3 A-level points

Notes: V = violent, NV = non-violent. Parameter estimates are listed in Web Appendix Table WA5

VI.   Conclusion

This paper investigates the effects of bullying in secondary school on later academic and labour market outcomes. We do this by exploiting a rich conditioning set of observables, and using a range of estimation methods: OLS, matching and weighting. The data come from a large high-quality cohort study in England, LSYPE, linked with administrative data on education records. Our empirical findings show that school bullying has negative consequences for short-term academic outcomes and persists to have adverse *long-term* effects – the strongest effects are on mental health, but we also find effects on unemployment and income measured at 25 years. We also find some evidence that type and intensity of bullying matters. For example, for the gold standard academic progression criterion of having 5 + GCSEs at age 16, the results suggests that it is not a higher degree of violent bullying alone that matters, it is the combination of both violent and non-violent bullying that matters. Similarly, in the case of mental (ill) health, it is again the combination of violent and non-violent bullying that generates much of the effect. The effect of low (medium, high) intensity non-violent bulling is a 0.6 (0.7, 0.9) of an SD increase in mental ill health, but add low (medium, high) intensity violent crime and the combined treatment effects is 0.7 (0.9, 1.4) of an SD.

We conduct a comprehensive battery of sensitivity tests to explore our main identifying and estimation assumptions. The results of this indicate that it is unlikely that our effects are entirely driven by selection on unobserved variables. A cautious interpretation of the results is that any of our effect sizes could potentially be reduced, but not eliminated, by unobserved selection. Even in this scenario, the estimated effects remain large enough to be of substantive importance. The most robust effects are on mental ill-health and unemployment. Being bullied exerts long run adverse effects of children's life outcomes. Based on our analyses, we feel confident that this finding is not an artefact of a particular estimation or identification assumption. If we take the mental health effects alone the costs associated with such an increase would be important enough to justify greater effort in reducing bullying. This conclusion is in line with existing literature, which consistently findings large detrimental effects of school bullying on mental health, across various identification strategies and measures of mental health (Sarzosa and Urzua, 2020; Rees et. al, 2020). Thus, the results have relevance for policy. Our results further suggest that low levels of non-violent bullying have modest effects, but higher intensity bullying has much larger effects. These findings suggest that the long run consequences of bullying should not be underestimated, and perhaps policy should be targeted more heavily on the extreme cases of violent and persistent cases.

There are several important ways that this research might, data permitting, be extended to broaden the reach of policy relevance. First, we do not analyse cyber-bullying. The children in the LSYPE data used here do not report explicitly on cyber-bullying. However, child and main parent reports of cyber-bullying are reported in LSYPE2 that was collected for a cohort 10 years later than LSYPE, when smartphone use had become more prevalent among young people[21]. Second, it would be useful to

---

[21]10% of pupils reported being sent offensive or hurtful texts, emails or online posts.

explore workplace bullying and its effects. The age 25 follow-up does contain contemporaneous bullying information and it shows a high correlation with bullying at age 14–16; however, the data from the planned further follow-up subsequent to age 25 year are not yet available. Third, we focus here on economic outcomes, and the work could be extended to a wider range of social outcomes.

Finally, the analysis relies on a selection on observables assumption. Although we explore the potential role of selection on unobservables via a battery of sensitivity analyses, which provides reassuring results, it would be useful to have confirmation from treatment effects based on exogenous variation and more objective measures of bullying. Such an analysis is not yet possible in our context because of data limitations and the lack of a suitable study design. We offer our results in the spirit of highlighting an important but difficult issue, for which we currently know relatively little about the magnitude and persistence of the consequences for young people.

Final Manuscript Received: May 2020

References

Anders, J. (2012). Using the Longitudinal Study of Young People in England for research into Higher Education access, DoQSS Working Papers 12–13, Department of Quantitative Social Science, University College London: Institute of Education.

Arseneault, L., Bowes, L.and Shakoor, S. (2010). ''Bullying victimization in youths and mental health problems: 'Much ado about nothing'?', *Psychological Medicine*, Vol. 40, pp. 717–729.

Bond, L., Carlin, J., Thomson, L., Rubin, K. and Patton, G. (2001). 'Does bullying cause emotional problems? A prospective study of young teenagers', *British Medical Journal*, Vol. 323, pp. 480–484.

Brown, K. (2018). Bullying: A review of the evidence, education policy institute blog post, https://epi.org.uk/publications-and-research/bullying-a-review-of-the-evidence/.

Brown, S. and Taylor, K. (2008). 'Bullying, education and earnings: Evidence from national child development study', *Economics of Education Review*, Vol. 27, pp. 387–401.

Carrell, S. and Hoekstra, M. (2010). 'Externalities in the classroom: How children exposed to domestic violence affect everyone's kids', *American Economic Journal: Applied Economics*, Vol. 2, pp. 211–228.

Cattaneo, M. (2010). 'Efficient semiparametric estimation of multi-valued treatment effects under ignorability', *Journal of Econometrics*, Vol. 155, pp. 138–154.

Department for Communities and Local Government (2015). *The English Index of Multiple Deprivation (IMD) 2015—Guidance*, HMSO, London.

Due, P., Holstein, B. E., Lynch, J., Diderichsen, F., Gabhain, S. N., Scheidt, P., Currie, C. and the Health Behaviour in School-Aged Children Bullying Working Group (2005). 'Bullying and symptoms among school-aged children: International comparative cross-sectional study in 28 countries', *European Journal of Public Health*, Vol. 15, pp. 128–132.

Eriksen, T., Nielsen, H. and Simonsen, M. (2014). 'Bullying in Elementary School', *Journal of Human Resources*, Vol. 49, pp. 839–871.

Fiorini, M. and Keane, P. (2014). 'How the allocation of children's time affects cognitive and non-cognitive development?', *Journal of Labour Economics*, Vol. 32, pp. 787–836.

Ford, R., King, T., Priest, N. and Kavanagh, A. (2017). 'Bullying and mental health and suicidal behaviour among 14-to 15-year-olds in a representative sample of Australian children', *Australian and New Zealand Journal of Psychiatry*, Vol. 51, pp. 897–908.

Ichino, A., Mealli, F. and Nannicini, T. (2008). 'From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity?', *Journal of Applied Econometrics*, Vol. 23, pp. 305–327.

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

Ladd, G., Ettekal, I.. and Kochenderfel-Ladd, B. (2017). 'Peer Victimization trajectories from kindergarten through high school: Different pathways for children's school engagement and achievement?', *Journal of Educational Psychology*, Vol. 109, pp. 826–841.

Mendolia, S. and Walker, I. (2015). 'Youth unemployment and personality traits', *IZA Journal of Labour Economics*, Vol. 4, pp. 1–26.

Nannicini, T. (2007). 'Simulation-based sensitivity analysis for matching estimators', *Stata Journal*, Vol. 7, pp. 334–350.

OECD (2017). *How Much of a Problem is Bullying at School?* PISA in Focus No. 74. OECD Publishing, Paris.

Olweus, D. (1997). 'Bully/victim problems in school: Facts and intervention', *European Journal of Psychology of Education*, Vol. 12, pp. 495–510.

Oster, E. (2019). 'Unobservable selection and coefficient stability: theory and evidence', *Journal of Business & Economic Statistics*, Vol. 37, pp. 187–204.

Rees, D. I., Sabia, J. J. and Kumpas, G. (2020). Anti-Bullying Laws and Suicidal Behaviors among Teenagers (No. w26777). National Bureau of Economic Research.

Ren, D. and Voelkel, J. G. (2017). 'Bullying', in Zeigler-Hill, V. and Shackelford, T. (eds), *Encyclopedia of Personality and Individual Differences*, Berlin: Springer.

Rotter, J. B. (1966). 'Generalized expectancies for internal versus external control of reinforcement', *Psychological Monographs*, Vol. 80, pp. 1–28.

Sarzosa, M. and Urzua, S. (2020). Bullying in Teenagers, The Role of Cognitive and Non-Cognitive Skills, Quantitative Economics (accepted).

Sharp, S. (1995). 'How much does bullying hurt? The effects of bullying on the personal wellbeing and educational progress of secondary aged students', *Educational and Child Psychology*, Vol. 12, pp. 81–88.

STATA Corp. (2017). *Stata Statistical Software: Release 15*, StataCorp, College Station, TX.

Thompson, F. and Smith, P. (2010). The Use of Effectiveness of Anti-Bullying Strategies in Schools, Research Report DFE-RR098.

Ttofi, M. and Farrington, D. (2011). 'Effectiveness of school-based programs to reduce bullying: systematic and metaanalytic review', *Journal of Experimental Criminology*, Vol. 7, pp. 27–56.

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2017a). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. [data collection]. 3rd Edition. UK Data Service. SN: 7104, http://doi.org/10.5255/UKDA-SN-7104-3.

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2017b). Next Steps: Sweeps 1-8, 2004-2016, [data collection]. 14th Edition. UK Data Service. SN: 5545, http://doi.org/10.5255/UKDA-SN-5545-6.

Woods, S. and Wolke, D. (2004). 'Direct and relational bullying among primary school children and academic achievement', *Journal of School Psychology*, Vol. 42, pp. 135–155.

Wooldridge, J. M. (2007). 'Inverse probability weighted estimation for general missing data problems', *Journal of Econometrics*, Vol. 141, pp. 1281–1301.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd edn, MIT Press, Cambridge, MA.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Web Table WA1. Attrition by bullying status.
Web Table WA2. Multivariate distance matching.

Web Table WA3. Nearest neighbour propensity score matching.

Web Table WA4. OLS estimates of the effects of bullying (binary and continuous) on boys vs girls.

Web Table WA5. Full IPWRA results.

Web Figure WA1. Histogram showing common support and balance of the matched sample. All observations are on the common support.

Web Figure WA2. Plot summarizing the balance statistics comparing the unmatched and matched sample.

Figure WA3. Outcome means by type of bullying and level of bullying.