EdWorkingPaper No. 24-1088

Causal Mechanisms of Relative Age Effects on Adolescent Risky Behaviours

Luca Fumarco
Masaryk University, IZA, GLO

Francesco Principe
University of Bergamo

We investigate the mechanisms by which a student's age relative to classmates (i.e., relative age) influences risky health behaviors among European adolescents. Using a two-stage least squares approach, we show that relatively young students are more prone to engage in risky behaviors. These results hold after controlling for absolute age, country fixed effects, and birth season effects. In the second part of the paper, we conduct two sets of analyses on possible mechanisms. First, causal mediation analyses reveal that students' perceived academic performance is the primary mediator. Second, additional analyses suggest that perceptions of substance risks and peer usage prevalence may also play a significant role.

VERSION: November 2024

Suggested citation: Fumarco, Luca, and Francesco Principe. (2024). Causal Mechanisms of Relative Age Effects on Adolescent Risky Behaviours. (EdWorkingPaper: 24 -1088). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/d21r-ve74

Causal Mechanisms of Relative Age Effects on Adolescent Risky Behaviours

Luca Fumarco and Francesco Principe*

November 4, 2024

Abstract

We investigate the mechanisms by which a student's age relative to classmates (i.e., relative age) influences risky health behaviors among European adolescents. Using a two-stage least squares approach, we show that relatively young students are more prone to engage in risky behaviors. These results hold after controlling for absolute age, country fixed effects, and birth season effects. In the second part of the paper, we conduct two sets of analyses on possible mechanisms. First, causal mediation analyses reveal that students' perceived academic performance is the primary mediator. Second, additional analyses suggest that perceptions of substance risks and peer usage prevalence may also play a significant role.

Keywords: Relative age; age effects; enrollment cutoffs.
JEL Codes: C26; I10; I12; I21.

*Fumarco: Masaryk University, GLO, IZA, Lipová 41a, 602 00 Brno-střed, luca.fumarco@econ.muni.cz. Principe: University of Bergamo, Via dei Caniana, 2, 24127 Bergamo BG, Italy, francesco.principe@unibg.it.

Acknowledgements: We are grateful to Patrick Button, Ingo Isphording, Lukas Laffers, and Giacomo Pasini, alongside participants at the AIES 2021 conference, HERB workshop 2023, 8th LEER conference, EALE 2023 conference and seminars at the JRC Ispra and University of Bergamo for helpful suggestions and feedback. Luca Fumarco acknowledges the generous support from the NPO "Systemic Risk Institute" number LX22NPO5101, funded by European Union - Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES) and from the CERGE-EI Foundation Teaching Fellowship Program. We acknowledge the HBSC study staff, which carries out the survey in collaboration with WHO/EURO. The International Coordinator is Candace Currie and the Data Bank Manager is Oddrun Samdal; a complete list of study coordinators, databank managers, researchers, and participating countries is available on the HBSC website. We acknowledge the ESPAD study staff; in particular, Sabrina Molinaro (ESPAD coordinator), João Matias (EMCDDA representative), Sharon Arpa, Pavla Chomynova, Ludwig Kraus, Ola Ekholm, Julian Strizek, and Loredana Fortunato.

1 Introduction

There is a vibrant public debate on within-class age differences, especially in the US, which has reinvigorated in the aftermath of the COVID-19 pandemic.[1] These within-class age differences (henceforth referred to as relative age) are defined by students' ages compared to those of their classmates. Research has shown that relative age can have far-reaching effects on individuals' educational performance, as well as their mental and physical health. This study contributes to the existing literature by investigating the effects of relative age on risky behaviors and, more importantly, exploring possible mechanisms.

Studies have showed that relative age is a leading cause of risky behaviors (Johansen, 2021; Argys & Rees, 2008). However, the literature on this topic has at least one limitation: the mechanisms behind these relative age effects on risky behaviors are still speculative. The limited empirically-driven discussion with respect to mechanisms limits policy makers' ability to design appropriate interventions. The aim of this paper is to fill this gap in the literature.

We use data from the international survey *Health Behaviour in School-Aged Children* (HBSC) on adolescents from more than thirty European countries. We start by employing the two-stage least squares methodology to address the endogeneity of relative age and find a series of robust results. Relatively young students are more likely to smoke tobacco and drink alcohol. This is the first paper to provide a causal estimate of relative age effects on marijuana use, showing that relatively young students are more likely to consume this substance. Unlike previous literature, our results suggest a less risky sexual life for relatively young students in terms of condom use. Finally, relatively old students are more likely to bully and to be involved in fights, although they do not differ from younger classmates in terms of cyberbullying. As an additional contribution to previous studies, we address both the endogeneity of relative age and disentangle its effect from that of absolute age.

[1] It is discussed in various news outlets, such as USNews and The Guardian, and at any governmental level, from school districts to intragovernmental institutions, such as UNESCO.

In the second part of the study, we investigate mechanisms in two ways. First, we contribute to the small but growing applied economics literature that uses causal mediation analyses, with Dippel et al. (2022) being one of the most prominent studies. In our analyses, we explore the mediating effects of academic self-concept, well-being, self-esteem, autonomy with respect to risky health behaviors, and peer support.

This investigation on the role of the mediators returns policy relevant results. We find that academic self-concept, as proxied by students' perception of their own school performance in relation to their classmates, is the most important indirect channel through which relative age affects risky health behaviors. This finding implies that simple solutions could be adopted to mitigate risky behaviors in adolescence. For example, the usage of "age allowances" to age-adjust the grades of the youngest students in the classroom could have the potential to mitigate public health issues caused by risky behaviors in adolescence (Peña, 2022).

These results could be causally interpreted, as we test the "sequential ignorability" assumption (Dippel et al., 2022; Celli, 2022; Dippel et al., 2019; Imai et al., 2010). This is an overidentification test conducted on the instruments of the mediators–which are added to the main analyses and are endogenous. Results from this test suggest the validity of the instruments of the mediators. Typically, mediators endogeneity cannot be addressed, which leads to the interpretation of the results in association terms (Celli, 2022).[2]

Second, we further study mechanisms behind relative age effects on consumption of addictive substances. In particular, this is the first study to investigate whether relative age is associated with perceived risk of consumption and perceived peers' prevalence of consumption. We find two results. First, relatively young students perceive lower risks associated with the consumption of addictive substances. Second, compared to their older classmates, they perceive that the prevalence of consumption of these substances is higher among their peers. The combination of these results suggests that the effect of relative age on addic-

[2]This issue is intuitively discussed in Datacolada.

tive substances consumption could be countered with informational campaigns, specifically targeting the youngest students in a cohort.

By investigating relative age effects on adolescents' risky behaviors and its mechanisms, we also contribute to the literature on the role of the educational system and peers in determining individual risky behaviors. For example, Card & Giuliano (2013) find significant peer effects in adolescents' sexual initiation, smoking, marijuana use, and truancy. These effects are greater for females. Elsner & Isphording (2018) find a negative effect of high school students' rank on the likelihood of smoking, drinking, having unprotected sex, and engaging in physical fights. Reynoso & Rossi (2019) exploit a natural experiment in Buenos Aires and show a relationship between attending high school at night and the probability of having unsafe sex or consuming substances, finding evidence on the absence of parental supervision as a key mechanism.

The literature on the effects of relative age on risky behaviors is characterized by two additional limitations. First, relative and absolute age effects are usually not disentangled. So far, only a few studies have conducted this disentanglement (e.g., Peña & Duckworth (2018); Peña (2017); Cascio & Schanzenbach (2016); Black et al. (2011)); however, it is policy relevant (Dhuey, 2016). For example, while it is possible to mitigate absolute age effects by postponing students' school start–either through a policy or an individual decision from the child's parents (i.e., redshirting), this postponement would have an ambiguous impact on the effects of within-class age differences.[3] Because of the particular features of the HBSC survey, we can study relative age in isolation.

Second, since HBSC smallest sample unit is the class, we can interpret our results in terms of proper peer effects compared to the analysis of students within a grade or cohort through

[3]Someone may argue that redshirting might be effective in reducing relative age effects as well, but, in fact, it might be a zero-sum game. While a student who has been redshirted might benefit from it, those who are not redshirted would suffer from it. For example, a redshirted student X born 30 days before the school enrollment cutoff date would be 30 days older than the oldest student Y in her new class–assuming for simplicity this oldest student was born on the cutoff date. On the other hand, the youngest student Z in her new class should have been 364 days younger than the oldest regular student Y in that class–assuming for simplicity this youngest student was born one day before the cutoff. However, now this youngest student is 394 days younger than the new classmate X who has been redshirted.

classes, who might never interact in real life. There is broad consensus among scholars that classroom interactions are important determinants of risky health behaviors (Balestra et al., 2021).

The remainder of the paper is organized as follows. In the next section, we discuss the related literature on causal mediation analyses, relative age, and determinant of risky behaviors. Section 3 describes the data, whereas Section 4 presents the empirical strategy. Section 5 presents main results, while Section 6 discusses potential mechanisms. The final section provides a summary and concludes.

2 Literature Review

In this section, we discuss in greater detail the differences between this study and the small literature on relative age effects on risky behaviors. Johansen (2021) uses Danish register data and a fuzzy regression discontinuity design and find that women who were the youngest children in their cohort are more likely to abort and experience alcohol poisoning. Argys & Rees (2008) use data from the National Longitudinal Survey of Youth and find that female students with younger peers use substances more frequently than female students with older peers, while there is no equivalent effect on male students.

Argys & Rees (2008) does not address endogeneity of relative age; however, differently from Johansen (2021), it effectively disentangles relative from absolute age effects. Relative age in isolation is a peer effect; it is an externality generated by being older than classmates–conditional on absolute age–and it includes effects such as increased self-esteem, popularity, and aspiration.[4]

[4]While the relative age effect is the effect of age differences between classmates, there are three main absolute age effects in the literature. Age-at-school start effect, that is, the effect of the age at which students start school, and this is what the literature usually focuses on. Age-at-outcome effect, that is, the effect of the age at which the outcome was measured (or the survey was conducted). Time-in-school effect, that is, the effect of the time spent in school. All these different but related age effects cannot be typically disentangled. Due to the features of the dataset at hand, we are able to isolate relative age from the other three factors. While age at school start–which is given by country-wise regulations, is captured by country fixed-effects, we are still not able to disaggregate age-at-outcome from time-in-school, but this is not

There are four more studies that focus on a subset of risky health behaviors. Routon & Walker (2022) find mild evidence that the old students in a grade drink more alcohol. Bahrs & Schumann (2020) find that adults in their mid-thirties, who were the old students in their school grade, are less likely to smoke. Ballatore et al. (2020) and Mühlenweg (2010) find that the young students in their school grade are more likely to be victimized. While these studies address the endogeneity of relative age, they do not effectively disentangle relative age from absolute age and investigate a limited set of risky health behaviors.

None of the above studies can effectively investigate mechanisms through which relative age affects risky behaviors. Page et al. (2017) is an exception with this respect. They conduct lab experiments in Australia and find two types of results: (i) negative relative age effects on risky attitudes, concerning events out of people's control (i.e., BART test, where participants guess the number of pumps to a balloon before it explodes), and (ii) positive effects on risk-seeking attitudes, where the perceived level of risk depends upon respondents' experience (e.g., risky behaviors while driving), which might be given by greater self-confidence.

Compared to the above studies, we address endogeneity of relative age and separate its effect from that of absolute age; most importantly, we additionally focus on mechanisms.

3 Data and Variables

We use data from the "Health Behaviour in School-Aged Children" (HBSC), a cross-national survey on adolescents' health and well-being, administered in schools every four years. In our analysis, we use five consecutive waves: 2001/2, 2005/6, 2009/10, 2013/14, and 2017/18. We exclude countries for which we do not have certain information on the school enrollment cutoff date (e.g., Albania, Arzebaijan, Serbia), for which the cutoff date fell in the middle of the month–because we do not have the precise day of birth (e.g., Portugal, Roumania,

to focus of our paper. Typical studies that use RDD to investigate age at school start provide an estimate which includes relative age: two students born around the cutoff have about a similar absolute age when they are observed, but they started school one year apart (one of them spent also one additional year in the kindergarten) *and* they are on the opposite relative age spectrum within their school class.

Israel),[5] or that adopted multiple cutoffs within country (e.g., Switzerland, Germany, US, Canada)–because we do not have information on the region or state of the school. The final sample is composed of more than 600,000 students from 32 European countries with vastly different characteristics, which qualifies our study as a multi-country study with great external validity. This is a rarity in the literature on relative age. Table B.1 in the Appendix provides the number of observations by country and by wave, with country-specific cutoff dates.

The sampling unit at the class level, and the large variation in both absolute age and country-specific cutoff dates, makes the HBSC data particularly attractive. Respondents' age ranges between 10.5 and 17.5 years, and cutoff dates–as well as age at school entry–vary across countries. These sources of variation allow us to disentangle relative age from both absolute age and season-of-birth effects, as discussed below in greater detail.

Outcome variables. We study the probability of having conducted various risky behaviors: *(i)* Early smoking, a dummy variable which equals one if the student has already smoked once by 13 years of age; *(ii)* Smoking, a dummy variable which equals one if the student currently smokes; *(iii)* Early drinking, a dummy variable which equals one if the student has already drunk once by 13 years of age; *(iv)* Ever drunk, a dummy variable which equals one if the student has been drunk at least once in life; *(v)* Early Marijuana, a dummy variable which equals one if the student has smoked marijuana at least once by 13 years of age; *(vi)* Ever smoked marijuana, a dummy variable which equals one if the student has smoked marijuana at least once in life; *(vii)* Ever sex, a dummy variable which equals one if the student has had sex at least once in life; *(iix)* Unprotected sex, a dummy variable which equals one if the student has had sex without a condom during the last intercourse; *(ix)* Fight, a dummy variable which equals one if the student was involved in at least one fight in the past year; *(x)* Bully, a dummy variable which equals one if the student has bullied

[5] This missing information precludes us from using the regression discontinuity design.

Table 1: Risky health behaviors.

Variable	N	Mean
Early smoking	308,506	0.241
Smoking	497,116	0.793
Early drinking	307,860	0.380
Ever drunk	480,690	0.255
Early marijuana	49,857	0.032
Ever marijuana	247,255	0.139
Ever sex	206,845	0.205
Unprotected sex	45,766	0.369
Fight	598,799	0.366
Bully	603,359	0.273
Bullied	603,444	0.282
Cyberbullied	242,366	0.143

Note: All these risky behaviors are dummy variables.

someone at least once during the past two months; *(xi)* Been bullied, a dummy variable which equals one if the student has been bullied at least once during the past two months; *(xii)* Cyberbullied, a dummy variable which equals one if the student has been cyberbullied (either with pictures or messages) at least once in life.[6]

To the best of our knowledge, relative age effects on outcomes *(ix)* and *(xii)* have not been previously studied.[7] However, they provide useful insights. Participation in fights is a more precise indicator of physical violence than bullying; this is important, because physical violence between underage people can lead to legal consequences in some places.[8] Cyberbulling is on the rise since the increase in social media platform usage.[9]

Table 1 reports the number of observations and means per risky health behavior.

[6]Sample sizes varies across analyses of these outcomes. Questions on *(i)* to *(iix)* are asked only in classes where expected students are at least 15 years old at the moment of the survey. Questions on *(i)*, *(ii)*, *(iii)*, and *(iv)* are not asked in wave 2018. Outcome *(v)* is based on observations from wave 2014 only. Outcome *(xii)* is based on observations from waves 2014 and 2018 only.

[7]Similarly, we are not aware on any study on age at school start on these two outcomes.

[8]For example, in European countries, and in some US states, such as in Illinois, where students can be accused of assault and battery.

[9]This was particularly true during the worse time of the COVID-19 pandemic, as discussed by the UN; however, this most recent period is not covered by our data.

Relative age. Relative age, RA_{ic}, is measured as the difference between the age of the student i's in class c, AGE_{ic}, and the age of the oldest regular student I in class c, AGE_{Ic}; thus, this measure varied by class and its increase implies that student i is relatively older. By "regular student" we mean that the student is in the right class based on their age and on the country cutoff date. Note that this measure of relative age, which varies by class, can be constructed because the primary sample unit in the HBSC survey is the class.[10] More formally, relative age is constructed as in Equation 1:

$$RA_{ic} = AGE_{ic} - max(AGE_{Ic} | I \in R_c) \tag{1}$$

For regular students i in class c, $i \in R_c$, this measure ranges between zero (i.e., student i is the oldest regular student in the class) and -1 (i.e., there is almost one year difference between student i and the oldest regular student in the class).[11]

There are several reasons why relative age, RA_i, is likely to be endogenous. Parents may greenshirt (i.e., expedite school entry) their children or have them skip a grade. At the opposite, parents may redshirt (i.e., delay school entry) their children or teachers may decide to retain them. On one hand, children who begin school early or skip a grade are likely to be born right after the cutoff date, and thus to be perceived as being especially skilled compared to (younger) children in the same school cohort; on the other hand, children who begin school later or are retained are likely to be born right before the cutoff date, and thus to be perceived as having developmental problems with a higher frequency than (older) children in the same school cohort (Peña, 2017; Schwandt & Wuppermann, 2016; Sprietsma, 2010; Bedard & Dhuey, 2006). On top of that, in some countries the decision to redshirt may vary by parents' socio-economic status (SES) (Bedard & Dhuey, 2006). We address concerns of endogenous selection into treatment by implementing an instrumental variable

[10]The class size ranges between a minimum of 8 and a maximum of 31 students, with median and mean of 18 students, with a standard deviation of about 5.

[11]It is *almost* one year, because exactly one year would mean that student i was born on the same day, but in the next academic year.

Table 2: Descriptive statistics for (expected) relative age, (expected) absolute age, control variables.

Variable	N	Mean	SD	Min	Max
Relative age	597,327	-0.306	0.454	-5.750	5.167
Absolute age	616,973	0.001	1.646	-3.703	3.463
Expected relative age	616,973	5.529	3.373	0	11
Expected absolute age	616,973	13.521	1.635	10.250	17
Female	616,973	0.508			
Parents	596,387	0.760			
SES: Low	616,973	0.367			
SES: Medium	616,973	0.229			
SES: High	616,973	0.403			

Note: SES: Low is the reference dummy for the student's family socio-economic status. Female, Parents, and the SES variables are dichotomous variables; thus, minimum (min) and maximum (max) value, as well as standard deviation (SD), are not reported. Analyses additionally include vectors for wave, country and season of birth.

strategy, where the instrument is expected relative age.

Table 2 reports observations, means, standard deviations, minimum and maximum values for relative age, and the other independent variables. Statistics on minimum and maximum values suggest the presence of outliers for relative and absolute age. Regular students represent 83% of the sample, while non-regular students represent 15% of the students. The percentage of non-regular students with with extreme values is relatively low, being 5%.[12]

Expected relative age. Expected relative age, ERA_{iCOU}, is the month of birth of student i within the academic year of country COU and measures the number of integer months between the cutoff date and the academic month of birth of student i, as if this student entered school when they were supposed to.[13] It is a discrete variable that ranges

[12]We define extreme values of relative age as being either below -2 or above one. Thus, we exclude students between zero and one and between -1 and -2, that we consider them as non-extreme values. In fact, students between zero and one do not fall within the expected range, but they are retained or redshirted students. Students between -1 and -2 are greenshirted students or students who skipped a grade. Students with extreme relative age values are usually older than expected and could be students who have been retained several times and/or with actual learning disadvantages (i.e., beyond the ones simply due to a lower relative age); these students represent the 4.7% of the sample. Students who are younger than expected and have extreme relative age values represent 0.2% of the sample. With the data at hand, we cannot tell what percentage of these extreme relative age values are due to measurement error.

[13]This instrument, or similar versions, have already been used in the literature in Fumarco et al. (2020);

between zero and 11, with zero being the reference month that starts with the cutoff date and 11 the month that precedes the cutoff date. Therefore, ERA does not vary by class, but it varies by country due to the country variation in cutoff dates. A similar instrument is used in studies based on fuzzy RDDs.

Our analyses use a disaggregated version of this variable, as suggested in Angrist & Pischke (2008), that is, a vector of dummies for the academic month. Let us make a practical example to see how it looks. Consider a student born in December in a country with September 1st as the cutoff date (e.g., Luxembourg). Their ERA would be three, that is, she was born three months after the month starting with the cutoff date, for which ERA is set to zero. In the disaggregated version of this variable, the dummy for the academic month of birth three equals one, while dummies for other academic months of birth equal zero.

The disaggregation of expected relative age has three main benefits. First, this transformation increase the first-stage fit. Second, it allows us to test the validity of the instruments with a standard overidentifying restriction test; this is because there are 12 instruments for two endogenous variables. Third, it provides a straightforward way to test the "sequential ignorability" assumption of the mediation analyses. This test is composed of two overidentifying restriction tests, and, when both are passed, it suggests that the results from the mediation analyses can be causally interpreted. These two tests are on *(i)* the "ignorability of the mediators" and *(ii)* the "ignorability of the treatment" (i.e., the overidentification test from the main analyses on ERA), as called in Imai et al. (2010). Additional analyses are conducted with the discrete version of ERA, but the main results do not change.

Absolute age. Absolute age, AA_i, captures a mix of age at the time of survey-participation and time spent at school, which–similarly to RA–might depend on parents

Fumarco & Baert (2019); Page et al. (2019); Peña & Duckworth (2018); Peña (2017, 2020); Datar (2006). The latter studies measure ERA as the distance in non-integer years between student i's age–if they were a regular student–and the age of the hypothetically youngest (oldest) student in the class, who was born right before (right on) the cutoff date.

SES.[14] In any given class, students from high SES families might tend to be older (younger), if these families redshirt (greenshirt, depending on the country) their children more frequently. Thus, we instrument this variable with expected absolute age.

Expected absolute age. Expected absolute age EAA_i represents the absolute age that student i would have had in that class, if she was a regular student. It is the expected absolute age of students who participate in the same survey, live in the same country, attend the same classroom, and were born in the same quarter; thus, it is based on classmates age. This specification closely follows Peña & Duckworth (2018). As an alternative specification of this variable, we measure EAA as the median of classmates' (discrete) absolute age; in this way, all students in the same class have the same EAA, which ensures that EAA and ERA are orthogonal. However, results from the first and second stages with this version of EAA are virtually the same.

Control variables. We control for the standard set of demographic characteristics. First, we control for students' gender. This variable equals one for female students and zero for male students, the reference group.

Second, we control for whether the student lives with both parents. Note that a recent Danish study find that relative age may affect marriage stability (Landersø et al., 2020); however, balance tests suggest that relative age and family status were not related in our multi-country sample, see Section A.2. Moreover, we conduct robustness checks where we exclude this control variable and the main results remained unchanged, as expected by Frisch–Waugh–Lovell theorem.

Third, we account for the SES of students' families. SES is derived from multiple items according to the HBSC guidelines (C. Currie et al., 2008) and it is coded into three dummies for high, medium, and low SES, with the latter being the reference category. Note that a

[14]We explicitly focus on relative age effects, so the fact that we cannot separate the effects of age at the time of survey-participation and time spent at school does not matter for the sake of this study.

family's SES might be endogenous: it is the family's SES when the survey was conducted and, potentially, relative age might have influenced the family's SES since birth, through the influence on the mother's labour market outcomes (Landersø et al., 2020). We address this concern in a robustness check, where we exclude family SES from the control variables. These results are indistinguishable from the main results. Related to the above, SES might be suspected to affect RA and, more improtantly, ERA: some studies find that families with certain SESs might target different dates of delivery. To address this concern, we conduct balance tests on ERA; these results suggest that in Europe, on average, families with a different SES do not target different dates of birth. The related literature and these tests are discussed in greater detail in Section A.2.

Analyses also account for unobservable birthdate effects, known as "season-of-birth effects." The variable for season of birth is proxied by the month of birth within the calendar year (henceforth, calendar month) and ranges between zero (January, the reference month) and 11 (December). If left unaccounted for, season-of-birth effects could cause biased estimates. For instance, Bound & Jaeger (2000) explain that individuals born in wintertime more likely suffer from multiple health issues, such as mental disabilities and multiple sclerosis, while individuals born in Spring are more likely to be shy. Thus, season-of-birth effects capture calendar-period specific effects on health outcomes that do not depend on maturity differences, but that might cause differences between students born in different periods of the year. Other studies providing evidence of season-of-birth effects are J. Currie & Schwandt (2013), which investigates the effect of time of conception on birth-weight, and Dustmann et al. (2022), Trudeau et al. (2016), and Wernerfelt et al. (2017), which study the effect of sunlight during pregnancy and infancy on asthma, birth-weight, and childhood obesity.

It is important to highlight that since our empirical setting leverages variation in cutoff dates, expected relative age does not overlap with the calendar month of birth. In Figure A.1, Section A.1, in the Appendix illustrates some examples.

Finally, the analyses account for wave and country fixed effects. Among other things, country fixed effects capture cross-country variation in statutory age at school entry and for country-specific law and social norms that might affect risky behaviors.[15]

Mediator variables. We investigate the mediation effect of five channels: *(a)* student's academic self-concept in relation to the classmates; *(b)* perceived well-being; *(c)* satisfaction with own body image; *(d)* time spent outside with friends in the evening; and *(e)* perceived support from other students.[16] Channel *(a)* is a measure of academic self-concept. It is given by the standardized output of the survey question asking students what the teachers might think about their school performance compared to classmates. The original variable ranges from zero (below average) to three (very good). Channel *(b)* is a proxy for well-being. It is an index created by applying the principal component analysis to the standardized output of survey questions asking students about their perceived life-satisfaction and self-rated health, which are closely associated variables. The original variable of life-satisfaction ranges from zero (worst possible life) to ten (best possible life). The original variable of self-rated health ranges from zero (poor) to three (excellent). Channel *(c)* is an index that measures self-image. It equals one if the student thinks their body is about the right size and zero otherwise (i.e., much or a bit too thin, much or a bit too fat). Channel *(d)* is a proxy for parents' supervision. It is the standardized output of the survey question asking students how many days per week they meet with friends after 8pm. The original variable ranges from zero to seven days per week. Channel *(e)* is a measure of peers' acceptance. It is an index created by

[15] We could not use class fixed effect because it would absorb all the variation in relative age, see Equation 1.

[16] This list of mediators is not meant to be complete. We choose them based on a survey of evidence from other fields, as they provide insights into various psychological and social factors that can contribute to the relationship between relative age and risky behaviors, helping to explain the underlying mechanisms at play. In fact, we chose these mediators because: (a) student's academic self-concept in relation to their classmates can influence risk-taking to assert status (Massey et al., 2008); (b) low levels of perceived well-being may lead to risk-seeking behaviors (Sorbring et al., 2014; Santini et al., 2020); (c) satisfaction with own body image can drive risky behaviors to conform to physical ideals (Granner et al., 2002; Wild et al., 2004; Woertman & Van den Brink, 2012; Gillen et al., 2006); (d) time spent outside with friends in the evening provides opportunities for risky behaviors due to increased independence and lack of parental supervision (Averett et al., 2011; Fletcher, 2012); and (e) perceived support from other students acts as a protective factor against risky behaviors through a sense of belonging and reduced likelihood of engaging in detrimental activities (Springer et al., 2006).

applying the principal component analysis to the output of survey questions on schoolmates enjoying spending time together, helping each other and being kind, and accepting other students as they are. The three original variables ranged from zero (strongly agree) to four (strongly disagree).

4 Empirical Strategy

To study the impact of a student's relative age on risky behaviors, we estimate relative age effects with a 2SLS.

The specification of the second stage is illustrated in Equation 2:

$$Y_i = \beta_0 + \beta_1 \widehat{RA}_i + \beta_2 \widehat{AA}_i + \mathbf{X_i}\zeta + \mathbf{FE_i}\delta + \mu_i \tag{2}$$

Index i indicates the individual. Y_i is one of the outcome variables measuring risky behaviors discussed in Section 3. $\widehat{RA}_i$ and $\widehat{AA}_i$ are the predicted values of relative and absolute age, respectively. $\mathbf{X_i}$ is a vector of covariates and includes dummies for gender, living with both parents, and family SES. $\mathbf{FE_i}$ is a vector of fixed effects for the calendar month of birth, survey wave, and country.

We define the first stage as in the following Equation 3:

$$Endogenous_i = \gamma_0 + \gamma_1 ERA_i + \gamma_2 EAA_i + \mathbf{X_i}\iota + \mathbf{FE_i}\phi + \nu_i \tag{3}$$

Where endogenous is either RA_i or AA_i. Before proceeding with the analyses, we conduct balance tests to verify that ERA is orthogonal with respect to observable characteristics. We find that ERA is balanced with respect to all observable characteristics, and in particular with parents' SES. These results on the unbiased nature of ERA, and thus on birth date exogeneity, extend to EAA as well. Section A.2 in the Appendix discusses these tests and the results in greater details.

For comparison sake, one could wonder about how relative age effects obtained with a 2SLS, while controlling for absolute age, compare to age at school start effects obtained with the similar fuzzy RDD. However, the dataset does not have information on date of birth, which is a fundamental piece of information for RDD analyses.[17]

5 Results

Table B.2 in the Appendix shows the two first-stage regressions. In Column (1), we observe that the effect of each dummy variable for ERA on observed RA increases *almost* monotonically. These estimates are highly statistically significant, and suggest the disaggregation into dummies could be more suitable than using the discrete version for RA. In Column (2), we observe that the correlation between EAA and observed AA is particularly high; this result is due to a large number of regular students, that is, students who are in the class where they were supposed to be, net of redshirting, greenshirting, retention, or grade skipping.

For brevity sake, Table 3 reports only the second-stage estimates of relative age effects for all outcomes. Full statistics, with estimates on control variables and 2SLS ancillary tests, are reported in the Appendix in Tables B.3, B.4, B.5.

Relative age coefficients can be interpreted as percentage changes in the probability to adopt a certain risky health behavior following an increase in relative age by one year (i.e., the hypothetical maximum age gap between regular students–about 12 months): in other words, its increase implies that the student is relatively older.

The results reveal a consistent pattern across outcomes. Relatively young students are more likely to engage in smoking and drinking; an increase by one year in relative age decrease the chances of smoking before turning 13 by about 4.4%, and the chances of currently smoking by 2.4%. An increase by one year in relative age decrease the chances of being drunk at least once before turning 13 by 6.9%, and of being drunk at least once in life by 5%. These

[17]The dataset does not even have information on the exact day of survey participation, which could allow to estimate the date of birth, coupled with age.

Table 3: Relative age on all outcomes.

	(1)	(2)	(3)	(4)
	Early smoking	Smoking	Early drinking	Ever drunk
Relative age	-0.044***	-0.024***	-0.069***	-0.050***
	(0.008)	(0.004)	(0.009)	(0.005)
N	285,742	464,318	285,190	448,722
	Early marijuana	Ever marijuana	Ever sex	Unprotected sex
Relative age	-0.014**	-0.022***	0.022**	0.055**
	(0.007)	(0.007)	(0.009)	(0.022)
N	45,247	230,232	194,101	42,477
	Bullied	Bully	Cyberbullied	Fight
Relative age	-0.012**	0.022***	0.002	0.041***
	(0.006)	(0.005)	(0.006)	(0.006)
N	566,168	566,196	224,131	561,632

Note: Second stage estimates from the 2SLS. Early smoking equals one if the student has already smoked once by 13 years of age. Smoking equals one if the student smokes at least once a week. Early drinking equals one if the student has drunk at least once by 13 years of age. Ever drunk equals one if the student has been drunk at least once in life. Early Marijuana equals one if the student has smoked marijuana at least once by 13 years of age. Ever smoked marijuana equals one if the student has smoked marijuana at least once in life. Ever sex equals one if the student has had sex at least once in life. Unprotected sex equals one if the student has had sex without a condom during the last intercourse. Fight equals one if the student was involved in at least one fight in the past year. Bully equals one if the student has bullied someone at least once during the past two months. Been bullied equals one if the student has been bullied at least once during the past two months. Cyberbullied equals one if the student has been cyberbullied at least once in life. ***, **, * indicate significance at 1%, 5% and 10%, respectively. Full statistics, with estimates on control variables and ancillary tests, are reported in the Appendix in Tables B.3, B.4, B.5

results are statistically significant at the 1% level. Relative to baseline probabilities, these are economically significant effects too; this is particularly true for having been drunk at least once in life, which is 20% of the average probability, see Table 1.

Results on smoking marijuana are consistent with those on smoking tobacco and drinking. An increase by one year in relative age decrease the chances of having smoked marijuana at least once before turning 13 by about 1.4%, and of having smoked marijuana at least once in life by 2.2%. Both these results are statistically significant at the 1% level. An increase by one year in relative age increase the chances to have had sex at least once in life by 2.2%, and the chances to have had sex without a condom during the last intercourse by 5.5%. These results are statistically significant at the 5% level. Relative to baseline probabilities, these are economically significant effects too; with the largest magnitude being that for having had sex without a condom in the last intercourse, which is 16% of the average probability, see Table 1.

Finally, an increase by one year in relative age decrease the chance of having been bullied by 1.2%, while it increases the probability of having bullied at least once in the previous two months by 2.2%. An increase in relative age by one year does not affect the probability of having been cyberbullied at least once in life, while it increases the probability of being involved in a physical fight by about 4.1%.

We conduct various robustness checks. We repeat the main analyses without disaggregating ERA; the caveat is that we cannot conduct the overidentification test without disaggregating ERA. We conduct additional analyses while omitting family SES, with different specifications of relative age,[18], with clustered standard errors at country level, with school fixed effects in lieu of country fixed effects, without students with extreme values as previously specified, and with an alternative version of EAA, that is, the median of classmates' (discrete) absolute age. By and large, the results are virtually identical.[19]

[18] The difference between own age and the average age in class (with or without student i's age) the minimum age of the youngest regular student in class, or within-class percentile age rank. In all these versions, the instrument is always ERA.

[19] These results will be added in additional tables to the Appendix.

Details on diagnostics are reported in Table B.3, B.4, and B.5 in the Appendix. For all analyses, under-identification tests reject the null hypothesis that the instruments are not correlated with the endogenous variable, while weak-identification tests suggests they are not weakly correlated.[20] Moreover, in most cases, overidentification tests fail to reject the null hypothesis that the instruments are uncorrelated with the second-stage error term. The overidentification test rejects the null hypothesis for: having smoked and drunk before turning thirteen, having ever smoked marijuana, having had unprotected sex, and having been bullied in the recent past. Thus, results on these outcomes should be considered with caution.

These main results are mostly statistically significant at conventional levels and their magnitude is economically significant. How do they compare to the closest studies? Results on smoking and drinking are in line with those from Elsner & Isphording (2018), who investigate the effect of performance on risky behavior among American adolescents. Results on smoking are in line with those from Bahrs & Schumann (2020).

The results on unprotected sex are opposite those of Johansen (2021). We should highlight three main differences between our study and Johansen (2021)'s. First, this result comes from a representative sample of adolescent students coming from most European countries, while Johansen (2021) use administrative data from Denmark. As Johansen (2021) writes, policies may affect differences between genders in Denmark, and, we think, they might play a key role in determining differences between countries too. Note that we include country fixed effects that capture differences in policies and/or norms. Second, in our analyses, boys and girls are considered together. In two additional analyses, we studied boys and girls separately. There, we observed no substantial difference between genders, with a point estimate of 0.061 for boys and 0.054 for girls, and statistical significance at 10% for both of estimates.[21] These results by gender should be considered with caution: the overidentification tests reject the null hypothesis for females. Third, we disentangle relative from absolute age

[20] For the latter, the F statistics are well beyond critical values suggested in Stock & Yogo (2005).
[21] These analyses will be added to the Appendix.

and we focus on within-class age differences; thus, we identify actual peer effects. Moreover, country fixed-effects capture differences in statutory age at school entry.

These results show why it is important to disentangle relative from absolute age. In this study, effects of absolute age go mostly in the opposite direction from those of relative age. For example, let us have a look at Tables B.3 and B.4 in the Appendix. The increase in absolute age by one year increases the chances of conducting risky health behaviors with respect to smoking and drinking as well as with respect to smocking marijuana and having had sex at least once in life, while it decrease the chances of having had sex without a condom during the last intercourse.

We conduct additional analyses at the country level in the Appendix, Section A.3. Here, we summarize the main findings. With some caveats, results at the country-level reflects those from the polled sample, suggesting great external validity of the main results. However, in some cases, results might differ due to country-specific aspects. Finally, RAEs on smoking and drinking behaviors might be mediated by adults' prevalence of consumption. These insights suggests that relatively younger students might be more prone to emulating adults' unhealthy drinking and smoking behaviors. More details are discussed in Section A.3.

6 Potential Mechanisms

In this section, we investigate potential mechanisms in two ways. First, we conduct causal mediation analyses and, second, we investigate whether relative age affects perceived risks of consuming addictive substances and perceived peers' prevalence of consumption of such substances.

6.1 Causal Mediation Analyses

We explore five channels through which relative age might affect risky behaviors: (a) student's academic self-concept in relation to her classmates; (b) perceived well-being; (c) satisfaction with own body image; (d) time spent outside with friends in the evening; and (e) perceived support from other students.

To assess the role of these five mechanisms, we conduct causal mediation analyses on five outcomes that pass the overidentification test in the main analysis and that, thus, are those that could pass the test of the sequential ignorability assumption. These outcomes are: (i) smoking, (ii) ever drunk, (iii) early marijuana, (iv) ever sex, (v) fight.[22]

Causal mediation analyses can be seen as a set of similarly unrelated regressions; for each outcome, there are 11 regressions. Regression one is Equation 2 and gives the "total effect" of relative age, C. Regressions two to six are expansions of Equation 2; the right-hand side of each of these regressions includes one of the five mediators. These regressions give the As, that is, the effects of the mediator on the outcome variable conditional on the treatment. These regressions have three endogenous variables: RA, AA, and one of the five mediators; these three variables are instrumented with the 11 dummies for ERA and with EAA. Regressions seven to eleven are similar to Equation 2, but the outcome is one of the five mediators. These regressions give the effect of relative age on the mediator, B. The "indirect effect", that is, the effect of relative age on risky health behaviors through the mediators, is found by multiplying A and B together. In this analysis, we focus on the percentage of the indirect effect out of the total effect, that is, [(A*B)/C]*100.

Different from past studies that use standard mediation analyses (e.g. Elsner & Isphording, 2017; Pagani et al., 2021), we use causal mediation analyses. That is, additionally to standard mediation analyses, we test the "sequential ignorability" assumption (Celli, 2022; Dippel et al., 2019; Imai et al., 2010). This test is actually composed of two complementary tests: *(i)* the ignorability of the mediators; and *(ii)* the ignorability of the treatment. Both tests are overidentifying restriction tests, and, if both are passed, results from the mediation analyses can be causally interpreted. The "sequential ignorability" assumption ought to be tested to be able causally interpret mediation analyses. The reason is that mediation anal-

[22] The overidentification test is passed also in the analyses on bully and cyberbullied, but we focus on fight because it is more policy relevant.

Table 4: Mediation analysis, % Indirect effect, for each mediator and each outcome.

	Smoking	Ever drunk	Early marijuana	Ever sex	Fight
	(1)	(2)	(3)	(4)	(5)
Academic self-concept	33.3%	25.7%	12%	31.6%	-28.6%
Well-being	12.2%	7.2%	3.3%	-8.5%	-9.3%
Body image	1.7%	1.9%	-0.2%	-0.1%	-3.1%
Evening out	-0.9%	-0.9%	0.8%	-2.2%	-1%
Students' support	3.5%	2.8%	6.2%	-8.8%	-9.3%

yses are conducted by adding a mediator that is typically endogenous to the set of control variables. [23]

Table 4 reports the percentage of the indirect effect out of the total effect. Tables B.6, B.7, B.8, B.9, and B.10 in the Appendix report the results in full. The tables in the Appendix report two additional statistics. First, the Hansen J statistics from the overidentification tests of the main analyses; within the causal mediation analysis framework, this statistic concerns the ignorability of the treatment. Second, the Hansen J statistics from the overidentification tests for the extended versions of Equation 2–one separate regression per mediator; within the causal mediation analysis framework, this statistic concerns the ignorability of the mediator.

There are three main results. First, academic self-concept, as measured by students' perception of their own school performance in relation to their classmates, is the most important indirect channel through which relative age affected risky health behaviors. The second most important channel is well-being, and then students' support. Second, we expected that frequent evenings spent with friends–which acts as a proxy for lower adults' supervision and thus greater autonomy with respect to health behaviors–would play an important role in channeling indirect effects, but we do not seem to observe this result. Third, self-image, as

[23] The updated version of the paper will include a different version of the tables in the section on causal mediation analyse, to more closely follow Dippel et al. (2022).

measured by students' satisfaction with their own body, does not seem to play an important mediation role either. All these results could be causally interpreted, since the test on the sequential ignorability assumption is passed; that is, the Hansen J statistics are quite large, as shown by Tables B.6, B.7, B.8, B.9, and B.10 in the Appendix.

There is one implication with respect to these results. While our school performance indicator does not allow us to investigate students' performance rank, our findings suggest that results on how performance rank affects risky behaviors might in fact be driven by within-class age differences. Said in other words, students rank affects risky behaviors and is determined by relative age. This result had to be largely expected, because of the abundant literature that shows the effect of relative age on students' performance; however, to the best of our knowledge, no study on students' rank considers classmates' peer effects caused by relative age as an explanatory mechanism.

6.2 Perceived Risk of Harming Themselves and Perceived Peers' Consumption Prevalence

To investigate in greater detail the potential mechanisms underlying the effect of relative age on the consumption of addictive substances, we conduct additional analyses on data from the "European Monitoring Centre for Drugs and Drug Addiction (ESPAD)" survey. This survey is on 15-16 year-old students, from 31 European countries/regions, and covers the period 1995-2015. The survey is conducted every four years since 1995; however, we could not use waves 1999 and 2003 because they do not include information on the month of birth. Similar to what we do in the main analyses with HBSC data, we exclude countries for which we do not have information on the cutoff date, the cutoff date falls in the middle of the month, or adopts multiple cutoffs across regions.[24] Table B.11 in the Appendix reports the

[24]Differently from the HBSC survey, the United Kingdom has one unique ESPAD survey, without specification on the state; since cutoff dates vary across its countries, we cannot use data from the UK. Moreover, differently from the HBSC survey, Germany is divided in two parts: Bayern and the rest of Germany, without specification on the state; since cutoff dates vary across states, we cannot use data from the rest of Germany, but we can use data from Bayern. Finally, observations from Russia, Albania and Moldova are excluded

full list of countries included in this analyses, with their respective cutoff, and the number of observations per wave. This table allows us to note the comparability of the two datasets: almost two-thirds of the countries in the ESPAD are also in the HBSC. These two surveys have additional points in common: the ESPAD's primary sampling unit is the class, many scientists are involved in the organization and study of both surveys, and ESPAD even coordinates some features with HBSC, such as the timing of the survey.[25]

For these analyses, we are unable to use the 2SLS and to conduct analyses at the class level due to two main reasons. First, with ESPAD data we cannot build a unique class identifier across wave, and we are not able to create it either; thus, analyses here are at the cohort level. Second, due to the sensitivity of the topics being dealt with in the ESPAD survey, the sample size is quite smaller than the HBSC one. Therefore, in these secondary analyses we focus on reduced-form effects, similar to Oosterbeek et al. (2021), and adopt the model specification in Equation 4:

$$Outcome_i = \gamma_0 + \gamma_1 ERA_i + \gamma_2 EAA_i + \mathbf{X_i}\zeta + \mathbf{FE_i}\phi + \nu_i \tag{4}$$

We regress the outcome on ERA and EAA, a vector of covariates that included dummies for gender, living with both parents, and SES. Moreover, there is a vector of fixed effects for the calendar month of birth, survey wave, and country. In these analyses, ERA is not disaggregated due to the smaller sample size (i.e., there would be fewer observations per academic month dummy) and ranges between zero and one, with one corresponding to being born in the month that starts with the cutoff. In this way, the interpretation of this coefficient is similar to that of RAEs in the main analyses. With this reduced form, we investigate seven

from the main analyses with HBSC data, because we do not know their cutoff with certainty, we do not have issues of statistical power, and we are concerned with precise estimates. Differently, in these analyses, we have concerns about statistical power, so we include these three countries despite the lack of certainty about the actual cutoff date being used. For sake of precaution, we conduct robustness checks without these three countries and obtain identical results, in terms of both magnitude and statistical significance; an exception is the result on smoke, which becomes statistically significant at 10% with a point estimate of -0.010. Notice that data on Russian students come from Moscow only–which is where a large percentage of Russians live. Data from Belgium, Wallonia, do not include information from month of birth, and thus it is not included.

[25] As explicitly stated in the ESPAD website.

outcomes that are divided into two categories.

Perceived risk of harming yourself while consuming addictive substances. Outcomes in this category equal one if the student answered the question "How much do you think people risk harming themselves (physically or in other ways), if they ..." with "more than slight risk." These are the five outcomes being investigated: (1) smoke an occasional cigarette; (2) have 4-5 drinks per week; (3) occasionally smoke marijuana; (4) try ecstasy; and (5) try amphetamine.

Perceived peers' frequency of consumption of addictive substances. Outcomes in this category equal one if the student answered the question "How many of your friends would you estimate ..." with "more than a few." These are the seven outcomes being investigated: (1) smoke; (2) drink alcohol; (3) get drunk; (4) smoke marijuana; (5) uses tranquilizers; (6) uses ecstasy; and (7) uses inhaler.

Tables B.12 in the Appendix reports observations and means for the outcomes, while Table B.13 reports observations, means, standard deviations, minimum and maximum values for expected relative and absolute age, and the other independent variables. Results on perceived risk are shown in Table 5.

This table shows that relatively old students are between 1.2 and 2.5 percentage points more likely to think that smoking tobacco and marijuana, drinking alcohol, and trying ecstasy could be risky activities. Although these point estimates are small, they would be reflected by large absolute numbers, out of the universe of adolescents in a country. There is no evidence of a reduced form association between relative age and thinking that trying amphetamines could be a risky behavior.

Results on perceived peers' frequency of consumption of addictive substances are shown in Table 6.

Table 5: Relative age effects on perceived risk of harming themselves - reduced form.

	Occasional cigarette (1)	4-5 drinks a week (2)	Occasional marijuana (3)	Try ecstasy (4)	Try amphetamine (5)
ERA	0.016***	0.012***	0.025***	0.012**	0.003
	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
N	121,572	120,480	113,768	106,354	105,922

Note: Outcome equals one if the answer to "How much do you think people risk harming themselves (physically or in other ways), if they ..." is more than slight risk. ERA is not disaggregated and ranges between zero and one, with one corresponding to being born in the month that starts with the cutoff. Control variables are expected absolute age, female, both parents at home, medium SES, high SES, fixed effects for the month of birth within the calendar year, country, and wave. Clustered standard errors at the level of the class in parentheses. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table 6: Relative age effects on perceived peers' consumption prevalence - reduced form.

	Smoke (1)	Drink (2)	Drunk (3)	Marijuana (4)	Tranquilizers (5)	Ecstasy (6)	Inhaler (7)
ERA	-0.007	-0.011***	-0.021***	-0.018***	0.003	-0.002	-0.004*
	(0.006)	(0.004)	(0.006)	(0.004)	(0.002)	(0.002)	(0.002)
N	95,474	95,292	90,850	95,026	93,136	95,263	94,345

Note: Outcome equals one if the answer to "How many of your friends would you estimate ..." is more than a few. ERA is not disaggregated and ranges between zero and one, with one corresponding to being born in the month that starts with the cutoff. Control variables are expected absolute age, female, both parents at home, medium SES, high SES, fixed effects for the month of birth within the calendar year, country, and wave. Clustered standard errors at the level of the class in parentheses. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

This table shows that relatively old students are between 1.1 and 2.1 percentage points less likely to think that the consumption of various addictive substances is frequent among their peers. There is no relative age effect on perceived peers' consumption prevalence of tranquilizers and ecstasy, while the effect on inhaler is also negative and statistically significant, but only at 10%.

Overall, these results suggest that relatively young students perceive that the consumption of addictive substances is less risky than what older peers think; moreover, they think that the consumption of these substances is more common among their friends than what older peers think. These results suggest that risky behaviors that entail the consumption of addictive substances might be due to relatively young students underestimating the risk associated with them. Moreover, they might either overestimate the actual consumption of addictive substances among their peers (i.e., they think that the consumption of these substances is more normal than it actually is) or, compared to their older classmates, they hang out more frequently with peers that consume these substances.

7 Conclusion

This paper investigates the effects of relative age within classroom on a comprehensive set of adolescents' risky health behaviors. In doing so, we contribute to two main branches of the economics literature: on relative age effects on risky behaviors and its determinants, and on the role of the educational systems and peers in determining individual risky behaviors.

Our results on relative age effects are very clear. Relatively young students are more likely to engage in risky behaviors such as smoking tobacco and marijuana and drinking alcohol. Differently, they are less likely to have sex without a condom. Moreover, relatively old students are more likely to be involved in fights and to be bullies, while relatively young students are more likely bullied, but they are not more likely to be cyberbullied.

Analyses at the country-level provide two insights. First, the main results seem to be highly "generalizable," at least in Europe. Second, as a preliminary insight into mechanisms, it seems that relatively younger students are more prone to have unhealthy drinking and smoking behaviors in those countries with a higher prevalence of smoking and alcoholism among adults.

Causal mediation analyses show that academic self-concept is the most important indirect channel through which relative age operate on risky health behaviors. This result implies that part of students' performance rank effect on risky behaviors might in turn be driven by within-class age differences. The second most important channels are well-being and classmates support.

Reduced-form results from secondary analyses suggest additional mechanisms. In particular, relatively young students underestimate the risk associated with the consumption of addictive substances. Also, they might either overestimate their actual peers' consumption of addictive substances or hang out more frequently with peers that consume these substances.

There are a number of policy implications stemming from these mechanisms results. The reduction of performance gaps may reduce health risky behaviors caused by relative age. Performance gaps could be reduced with "age allowances" that are conducted in a similar spirit to the golf handicapping system; this solution has a long tradition, although it is currently in use only in some places in England (Peña, 2022). Performance gaps could also be reduced through a targeted tutoring system; this system could be incentivized and focus on helping relatively young students with their studies. For example, echoing findings in Kraft & Falken (2021), higher education students could volunteer as tutors and help relatively young students with their courses. Effects on the consumption of addictive substances could be countered with informational campaigns specifically targeted to relatively younger students. Policy interventions to counter relative age effects on risky behaviors seems to be particularly pressing in countries with higher prevalence of adults smoking and alcoholism.

These studies on mechanisms help us understanding the policy implications of studies on relative age effects. These interventions might be relatively cheap and would increase the fairness of the educational system, while promising positive effects on public health and related costs. These interventions are justifiable based on equality concerns as well. The youngest students in a grade are experiencing a situation that resembles "indirect discrimination," as it is called in the EU and in some Anglo-Saxon countries' legal systems, that is, there is an apparently neutral rule that is applied to everybody (i.e., the cutoff that affects the age grouping in education) but that systematically and negatively affects people with a given demographic characteristic.[26]

To improve our knowledge of relative age effects, and in particular their policy implications, future studies should focus more systematically on the mechanisms behind them. Only a few studies have so far investigated these mechanisms, such as Page et al. (2019) and Page et al. (2017). For example, we do not know of any study that relates relative age to time-inconsistent preferences, which are known to cause agents to make worse consumption choices, also in terms of addiction and unhealthy diets (Gruber & Mullainathan, 2005; Read & Van Leeuwen, 1998).

[26] Greater details are provided, for example, in the Eurofund website, UK Statute Law Database website, and Australian Law Reform Commission website.

References

Alubaydli, O., & List, J. A. (2015). On the generalizability of experimental results in economics. *Handbook of Experimental Economic Methodology*.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton university press.

Argys, L. M., & Rees, D. I. (2008). Searching for peer group effects: A test of the contagion hypothesis. *Review of Economics and Statistics*, *90*(3), 442–458.

Averett, S. L., Argys, L. M., & Rees, D. I. (2011). Older siblings and adolescent risky behavior: does parenting play a role? *Journal of Population Economics*, *24*, 957–978.

Bahrs, M., & Schumann, M. (2020). Unlucky to be young? the long-term effects of school starting age on smoking behavior and health. *Journal of Population Economics*, *33*(2), 555–600.

Balestra, S., Sallin, A., & Wolter, S. C. (2021). High-ability influencers? the heterogeneous effects of gifted classmates. *Journal of Human Resources*, 0920–11170R1.

Ballatore, R. M., Paccagnella, M., & Tonello, M. (2020). Bullied because younger than my mates? the effect of age rank on victimisation at school. *Labour Economics*, *62*(C), S0927537119301083.

Bearak, J. M., Popinchalk, A., Beavin, C., Ganatra, B., Moller, A.-B., Tunçalp, Ö., & Alkema, L. (2022). Country-specific estimates of unintended pregnancy and abortion incidence: a global comparative analysis of levels in 2015–2019. *BMJ Global Health*, *7*(3), e007151.

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics*, *121*(4), 1437–1472.

Black, S., Devereux, P., & Salvanes, K. G. (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics*, *93*(2), 455-467.

Bo, H., & Galiani, S. (2021). Assessing external validity. *Research in Economics*, *75*(3), 274–285.

Bound, J., & Jaeger, D. A. (2000). Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? In *Research in labor economics.* Emerald Group Publishing Limited.

Buckles, K., & Hungerman, D. (2013). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics*, *95*(3), 711-724.

Card, D., & Giuliano, L. (2013). Peer effects and multiple equilibria in the risky behavior of friends. *Review of Economics and Statistics*, *95*(4), 1130–1149.

Cascio, E., & Schanzenbach, D. (2016). First in the class? age and the education production function. *Education Finance and Policy*, *11*(3), 225-250.

Celli, V. (2022). Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys*, *36*(1), 214–234.

Clarke, D., Oreffice, S., & Quintana-Domeque, C. (2019). The demand for season of birth. *Journal of Applied Econometrics*, *34*(5), 707–723.

Cook, P. J., & Kang, S. (2016). Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *American Economic Journal: Applied Economics*, *8*(1), 33–57.

Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: the development of the health behaviour in school-aged children (hbsc) family affluence scale. *Social Science & Medicine*, *66*(6), 1429–1436.

Currie, J., & Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. *Proceedings of the National Academy of Sciences*, *110*(30), 12265–12270.

Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, *25*(1), 43–62.

Dee, T. S., & Sievertsen, H. H. (2018). The gift of time? school starting age and mental health. *Health economics*, *27*(5), 781–802.

Dhuey, E. (2016). Age at school entry: How old is old enough? a child's age at school entry matters, and the implications of policy changes can have long-lasting effects. *IZA World of Labor*, *247*.

Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management*, *38*(3), 538–578.

Dhuey, E., & Lipscomb, S. (2010). Disabled or young? relative age and special education diagnoses in schools. *Economics of Education Review*, *29*(5), 857–872.

Dickert-Conlin, S., & Elder, T. (2010). Suburban legend: School cutoff dates and the timing of births. *Economics of Education Review*, *29*(5), 826–841.

Dippel, C., Gold, R., Heblich, S., & Pinto, R. (2019). Mediation analysis in iv settings with a single instrument. In *Working paper.*

Dippel, C., Gold, R., Heblich, S., & Pinto, R. (2022). The effect of trade on workers and voters. *The Economic Journal*, *132*(641), 199–217.

Dustmann, C., Sandner, M., & Schönberg, U. (2022). The effects of sun intensity during pregnancy and in the first 12 months of life on childhood obesity. *Journal of Human Resources*.

Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers. *Journal of Human Resources*, *44*(3), 641–683.

Elsner, B., & Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, *35*(3), 787–828.

Elsner, B., & Isphording, I. E. (2018). Rank, sex, drugs, and crime. *Journal of Human Resources*, *53*(2), 356–381.

Fletcher, J. M. (2012). Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, *25*, 1265–1286.

Fumarco, L., & Baert, S. (2019). Relative age effect on european adolescents' social network. *Journal of Economic Behavior & Organization*, *168*, 318–337.

Fumarco, L., Baert, S., & Sarracino, F. (2020). Younger, dissatisfied, and unhealthy–relative age in adolescence. *Economics & Human Biology*, *37*, 100858.

Furzer, J., Dhuey, E., & Laporte, A. (2022). Adhd misdiagnosis: Causes and mitigators. *Health Economics*, *31*(9), 1926–1953.

Gans, J. S., & Leigh, A. (2009). Born on the first of july: An (un) natural experiment in birth timing. *Journal of Public Economics*, *93*(1-2), 246–263.

Gillen, M. M., Lefkowitz, E. S., & Shearer, C. L. (2006). Does body image play a role in risky sexual behavior and attitudes? *Journal of Youth and Adolescence*, *35*, 230–242.

Granner, M. L., Black, D. R., & Abood, D. A. (2002). Levels of cigarette and alcohol use related to eating-disorder attitudes. *American Journal of Health Behavior*, *26*(1), 43–55.

Gruber, J. H., & Mullainathan, S. (2005). Do cigarette taxes make smokers happier. *The BE Journal of Economic Analysis & Policy*, *5*(1).

Huang, C., Zhang, S., & Zhao, Q. (2020). The early bird catches the worm? school entry cutoff and the timing of births. *Journal of Development Economics*, *143*(C), S0304387818312082.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71.

Johansen, E. R. (2021). Relative age for grade and adolescent risky health behavior. *Journal of Health Economics*, *76*, 102438.

Kraft, M. A., & Falken, G. T. (2021). A blueprint for scaling tutoring and mentoring across public schools. *AERA Open*, *7*, 23328584211042858.

Landersø, R. K., Nielsen, H. S., & Simonsen, M. (2017). School starting age and the crime-age profile. *Economic Journal*, *127*(602), 1096–1118.

Landersø, R. K., Nielsen, H. S., & Simonsen, M. (2020). Effects of school starting age on the family. *Journal of Human Resources*, *55*(4), 1258–1286.

List, J. A. (2020). *Non est disputandum de generalizability? a glimpse into the external validity trial* (Tech. Rep.). National Bureau of Economic Research.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale.* Crown Currency.

Massey, E. K., Gebhardt, W. A., & Garnefski, N. (2008). Adolescent goal content and pursuit: A review of the literature from the past 16 years. *Developmental Review*, *28*(4), 421–460.

Mühlenweg, A. (2010). Young and innocent: international evidence on age effects within grades on victimization in elementary school. *Economics Letters*, *109*(3), 157–160.

Oosterbeek, H., ter Meulen, S., & van der Klaauw, B. (2021). Long-term effects of school-starting-age rules. *Economics of Education Review*, *84*, 102144.

Pagani, L., Comi, S., & Origo, F. (2021). The effect of school rank on personality traits. *Journal of Human Resources*, *56*(4), 1187–1225.

Page, L., Sarkar, D., & Silva-Goncalves, J. (2017). The older the bolder: Does relative age among peers influence children's preference for competition? *Journal of Economic Psychology*, *63*, 43–81.

Page, L., Sarkar, D., & Silva-Goncalves, J. (2019). Long-lasting effects of relative age at school. *Journal of Economic Behavior & Organization*, *168*, 166–195.

Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, *56*, 152–176.

Peña, P. A. (2019). Relative age and incarceration: born on the wrong side of the calendar. *Education Economics*, *27*(6), 588–607.

Peña, P. A. (2020). Relative age and investment in human capital. *Economics of Education Review*, *78*, 102039.

Peña, P. A. (2022). End the birthday bias: Age allowances in high-stakes tests are a proven boost for fairness. *Education Next*, *22*(3), 22–29.

Peña, P. A., & Duckworth, A. L. (2018). The effects of relative and absolute age in the measurement of grit from 9th to 12th grade. *Economics of Education Review*, *66*, 183–190.

Read, D., & Van Leeuwen, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational Behavior and Human Decision Processes*, *76*(2), 189–205.

Reynoso, A., & Rossi, M. A. (2019). Teenage risky behavior and parental supervision: the unintended consequences of multiple shifts school systems. *Economic Inquiry*, *57*(2), 774–791.

Routon, P. W., & Walker, J. K. (2022). (more) party time in school? relative age and alcohol consumption. *Applied Economics*, 1–17.

Santini, Z. I., Meilstrup, C., Hinrichsen, C., Nielsen, L., Koyanagi, A., Koushede, V., ... Madsen, K. R. (2020). Associations between multiple leisure activities, mental health and substance use among adolescents in denmark: A nationwide cross-sectional study. *Frontiers in behavioral neuroscience*, *14*, 593340.

Schwandt, H., & Wuppermann, A. (2016). The youngest get the pill: Adhd misdiagnosis in germany, its regional correlates and international comparison. *Labour Economics*, *43*, 72–86.

Sorbring, E., Skoog, T., & Bohlin, M. (2014). Adolescent girls' and boys' well-being in relation to online and offline sexual and romantic activity. *CyberPsychology*, *8*(1).

Sprietsma, M. (2010). Effect of relative age in the first grade of primary school on long-term scholastic results: international comparative evidence using pisa 2003. *Education Economics*, *18*(1), 1–32.

Springer, A., Parcel, G., Baumler, E., & Ross, M. (2006). Supportive social relationships and adolescent health risk behavior among secondary school students in el salvador. *Social Science & Medicine*, *62*(7), 1628–1640.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80.

Trudeau, J., Conway, K. S., & Kutinova Menclova, A. (2016). Soaking up the sun: The role of sunshine in the production of infant health. *American Journal of Health Economics*, *2*(1), 1–40.

Wernerfelt, N., Slusky, D. J., & Zeckhauser, R. (2017). Second trimester sunlight and asthma: Evidence from two independent studies. *American Journal of Health Economics*, *3*(2), 227–253.

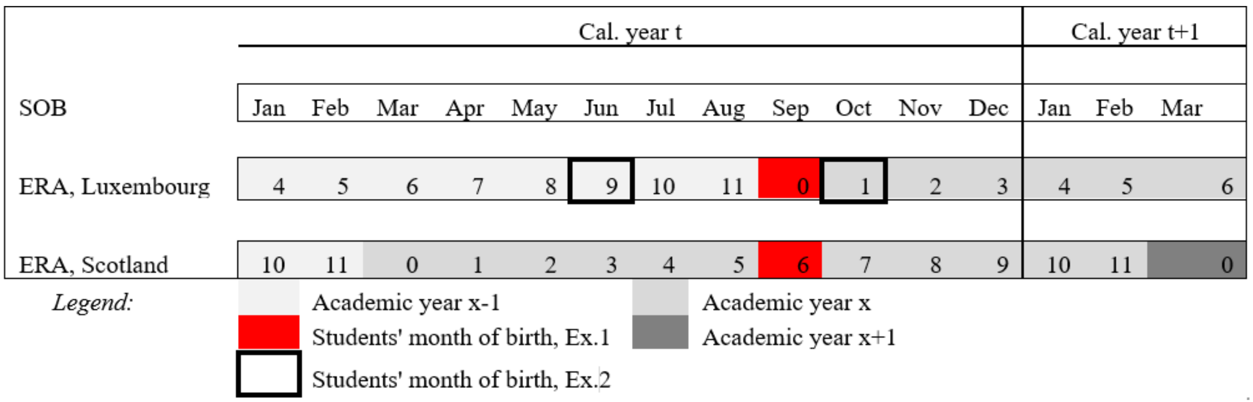
Wild, L. G., Flisher, A. J., Bhana, A., & Lombard, C. (2004). Substance abuse, suicidality, and self-esteem in south african adolescents. *Journal of drug education*, *34*(1), 1–17.

Woertman, L., & Van den Brink, F. (2012). Body image and female sexual functioning and behavior: A review. *Journal of sex research*, *49*(2-3), 184–211.

A Appendix

A.1 Expected relative age does not overlap with calendar month of birth

Figure A.1: Season of birth and expected relative age; example with Luxemburgish and Scottish students.



Note: This figure illustrates two examples. Red cells illustrate Example 1. Here, there are two children born in September of the same calendar year t–and thus in the same season of birth, but in countries with different cutoff dates: September 1st in Luxembourg and March 1st in Scotland. Thus, they have different Expected Relative Age (ERA). Cells with the thick boarder illustrate Example 2. The thick-boarder cell under June shows that this student's ERA is 9, being born on the 9th month of academic year x-1 (i.e., light grey cells, before September), in Luxembourg. However, this student was retained and is placed in academic year x (i.e., darker grey cells, after September), where the oldest regular student is born four months later, in the thick-boarder cell under October. Thus, the retained student born in June is about four months older than the oldest regular student in the same class, who was born in October.

Example 1. This example illustrates why season of birth differs from ERA, for two students born in the same year and month, but in two countries with a different cutoff date. There are two children born in September of calendar year t; one student was born in Luxembourg, where the cutoff is September 1st, and the other one was born in Scotland, where the cutoff is March 1st. The Luxemburgish student is among the oldest students in their class since they were born in the month that starts with the cutoff date, while the Scottish student was born six months after the cutoff date month. These two students' months of birth are identified by red cells.

Because of differences in cutoff dates, the correlation between ERA and season of birth was about 0.549. This value does not cause problems of multicollinearity in the first stage. Table A.1 reports the variance inflation factors (VIFs) for both ERA and season of birth from the first stage: most of them are below 4.5, while the mean factor of the entire first stage is 3.62; these results are reassuring since the rule of thumb suggests that multicollinearity could be a problem when VIF>10.

Example 2. This figure illustrates why RA differs from ERA, for two students born in the same year and country, but different month, and where one student was retained. Here, we consider a retained Luxemburgish student born in June, in the thick-boarder cell. They should be relatively young–being born in the ninth month of the academic year, but they are relatively old because of their retention. Assume that, in their current class, the oldest regular classmate (i.e., neither retained nor redshirted) was born in October, in the other thick-boarder cell. The retained student's ERA is nine because they were born nine months after the Luxemburgish cutoff date month; however, their (observed) RA is about 0.33, because they are about four months older than the actual oldest regular student in their class (i.e., the regular student born in October, in the thick-boarder cell).

Table A.1: Variance inflation factor of expected relative age and season of birth, from first stage.

Variables	VIF
ERA	
1	3.39
2	3.30
3	3.37
4	3.15
5	3.85
6	3.08
7	3.75
8	3.08
9	3.27
10	3.15
11	3.43
Season of birth	
1	4.03
2	3.95
3	4.18
4	3.51
5	4.62
6	3.81
7	4.46
8	3.35
9	4.08
10	3.69
11	4.18
Mean VIF	3.62

Note: VIFs' in this table, including the mean VIF, refer to the first stage.

A.2 Balance Tests

There is one underlying assumption for using ERAas instrument, which is in common with most of the other literature on relative age: birthdate has to be orthogonal to demographic variables (Dickert-Conlin & Elder, 2010).

Some studies find that families with different SES might target different dates of delivery. Evidence of this phenomenon, its magnitude, and its generalizability are a matter of debate. It has been found mostly in Anglo-Saxon countries (e.g., in the US (Clarke et al., 2019; Buckles & Hungerman, 2013) and in Australia (Gans & Leigh, 2009)) and in the vicinity of the cutoff date (e.g., as discussed in Dhuey & Lipscomb (2010), through cesarean section, which in many countries is restricted to medical motivations), and with limited economic significance elsewhere (e.g., in China (Huang et al., 2020)). In general, this relationship depends on local characteristics, such as norms and tax incentives, as explained in Dickert-Conlin & Elder (2010).

We address the concern on the possible non-orthogonality of birthdate and demographic characteristics in two distinct ways. First, we conduct robustness checks where we replicate the main analyses, but we exclude family SES from the set of control variables; the main results are unchanged. Second, we conduct joint orthogonality tests on the instrumental variable for relative age. In other words, we test the orthogonality of ERA with respect to observable demographic characteristics with a series of conditional balance tests, where we run OLS regressions of each demographic characteristic on the set of dummies for ERA and on fixed effects for country, wave, and season of birth. The results are reported in Table A.2 and they suggest that in Europe, on average, families with a different SES do not target different dates of birth. In other words, these results are reassuring as they suggest that ERA is randomly distributed with respect to observable characteristics. Most importantly, ERA is balanced with respect to parents' SES; this result rules out the possibility that parents tend to target certain birth dates depending on their SES–when we control for between-country differences. These results on the unbiased nature of ERA, and thus on birth date exogeneity, suggest two things: (i) EAA is exogenous too, and (ii) absolute age is exogenous and would not need instruments. This exogeneity of birthdate–and thus of absolute age–will be useful for conducting the mediation analyses à la Dippel et al. (2022).

Indeed, analyses where we do not instrument absolute age return results that are equivalent to the main ones. Also these results will be available in the specification curve in the updated version of the paper.

As a final note, we would like to stress that this result does not call into question results from the above literature on birthdate targeting, because these balance tests are conducted with country fixed-effects and thus are accounting for between-countries differences.

Table A.2: Conditional correlation between expected relative age and main control variables.

Variables	Female	Parents	SES: Low	SES: Medium	SES: High
	(1)	(2)	(3)	(4)	(5)
ERA 1	0.000	-0.003	-0.001	0.001	0.000
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
ERA 2	0.005	0.001	-0.000	-0.001	0.001
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
ERA 3	0.002	-0.002	-0.005	0.002	0.004
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
ERA 4	0.004	-0.001	-0.001	-0.002	0.003
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
ERA 5	0.001	-0.001	0.002	-0.000	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ERA 6	0.004	0.005	0.000	-0.003	0.002
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
ERA 7	0.010**	0.002	0.003	-0.001	-0.002
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
ERA 8	0.006	0.001	-0.000	0.002	-0.001
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
ERA 9	0.004	0.001	-0.002	0.003	-0.002
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
ERA 10	0.003	-0.005	0.003	0.001	-0.004
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
ERA 11	0.003	-0.004	0.004	-0.000	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N	616,973	596,387	616,973	616,973	616,973

Note: ERAs are dummies for academic month of birth, that is, expected relative age. SES is socio-economic status. Clustered standard errors at the class level in parentheses.

A.3 Heterogeneous Effects

Analyses at the country-level are useful to gain two insights. First, whether some countries drive the average relative age effect in the polled sample. With this respect, this exercise can be looked at as an extreme version of the leave-one-out exercise. Second, whether the average relative age effect is representative of the individual-countries' relative age effect. With this respect, these analyses can be looked at as an informal test on the external validity of the main results, by answering this question: does the same (quasi)experiment, repeated in different countries, lead to the same or similar result?

This external validity aspect is not usually investigated. However, it is relevant: effects on risky behaviors might change across countries depending on several factors, such as legal framework, social norms, and support to parenthood. While the issue of external validity is gaining traction in experimental economics and in studies that make use of other RCTs (Alubaydli & List, 2015; List, 2020; Bo & Galiani, 2021; List, 2022), it should be a concern in applied economics as well.[27] Depending on the outcome, this consistency might have important policy implications. For example, punishments for marijuana possession and their severity vary across countries—from being legal to entailing jail time. As a consequence, relative age effects on risky behaviors might impact youth criminal records in some countries, returning additional nuances to findings on the connection between relative age and youth criminality (Landersø et al., 2017; Peña, 2019; Cook & Kang, 2016).

These country-level analyses come with some caveats. Statistical power is reduced; moreover, the identification strategy is weaker: there is no within-country variation in cutoff and, thus, we cannot isolate relative age from season-of-birth effects. Other than that, these analyses replicate the main analyses, and they still allow us to separate relative from absolute age, because relative age varies at school-class level. We report results for each outcome, except for unprotected sex and early marijuana, due to the small sample size. Moreover, we exclude Greenland, because of the reduced sample size.

More concretely, we informally verify whether relative age effects vary with some countries characteristics. We look at how relative age effects on: (i) smocking marijuana vary with national legislations on incarceration for marijuana possession; (ii) alcohol consumption vary with the national adult men's alcoholism rate: (iii) smocking tobacco vary with the national adults' smoking prevalence; (iv) bullying vary with its prevalence within the HBSC dataset.

The results are illustrated in Figures A.2 and A.3.

[27] The seminal study from Bedard & Dhuey (2006) investigates relative age effects on school performance on individual countries with PISA data, which provides strong external validity. However, most following studies on various relative age effects use administrative data; these data limit researchers' ability to investigate the external validity of the results.

Before proceeding with the interpretation, note that country names are in bold if: (a) in that country, it is possible to be incarcerated for marijuana possession in that country; (b) in that country, adult men's alcoholism rate is greater than the median value among the countries in the sample (i.e., 4.5%); (c) in that country, adults' smoking prevalence (i.e., the proportion of the population aged 15 and older reporting daily smoking) is larger than the median value among the countries in the sample (i.e., 19%); (d) in that country, the percentage of adolescents who have recently victimized someone is greater than the median value among the countries participating in the sample (i.e., 26%); (e) in that country, the percentage of unintended pregnancies is greater than the median value among the countries in the sample (i.e., 12%).[28] These five variables represent proxies for the country's norms in terms of substances consumption, sexual freedom, and bullying.

Three insights are worth considering. First, the estimated RAEs on smoking tobacco and drinking alcohol are overall confirmed, while results on having ever had sex and smoked marijuana seem inconclusive and some extreme results might drive the main estimates in the polled sample. Results on fighting behaviors are overall confirmed, although we note lots of variance in cyberbulling. Second, results on having ever had sex for Italy are statistically significant and negative, opposite from the results in the polled sample and similarly to the results in Johansen (2021); this result suggest that indeed, country-specific aspects might matter. Third, RAEs on smoking and drinking behaviors might be mediated by adults' smoking behaviors. In fact, many countries with a negative effect on early and currently smoking tobacco are characterized by high adults' smoking prevalence. Same applies for having ever been drunk, for which the negative effect seems to be frequent in countries with high adults alcoholism prevalence. These insights suggests that relatively younger students might be more prone to emulating adults' unhealthy drinking and smoking behaviors. More-

[28] Information on incarceration for marijuana possession comes from the European Monitoring Centre for Drugs and Drug Addiction and from Wikipedia for Iceland and Ukraine. Information on the alcoholism rate in 2021 comes from Our World of Data and is elaborated by the Abbeycare Foundation. Information on smoking prevalence in 2019 comes from the World Health Organization and is elaborated by the Tobacco Atlas. Information on bullying rate comes from the HBSC dataset. Information on unintended pregnancy rate comes from Bearak et al. (2022).

over, negative (positive) effect on being bullied (being a bully) is less prevalent in countries where students are less used to bully each others

These results are purely descriptive and suggest an important aspect that is often neglected: while in most countries relative age effects on performance (Bedard & Dhuey, 2006; Datar, 2006; Elder & Lubotsky, 2009; Cascio & Schanzenbach, 2016; Peña, 2017; Dhuey et al., 2019; Peña, 2020) and on attention deficit and hyperactivity disorder (Furzer et al., 2022; Schwandt & Wuppermann, 2016; Dee & Sievertsen, 2018; Elder & Lubotsky, 2009) have a well-established direction (positive for performance and negative for ADHD), the direction of other relative age effects might depend on country-specific characteristics, such as norms, social welfare, and legislation. This insight goes hands in hands with those from the literature on relative age effects on juvenile crime, where country-specific factors, such as school-leaving requirements, seem to drive the results (Landersø et al., 2017; Cook & Kang, 2016).

Figure A.2: Relative age effects on smoking tobacco, drinking, smoking marijuana, and sex

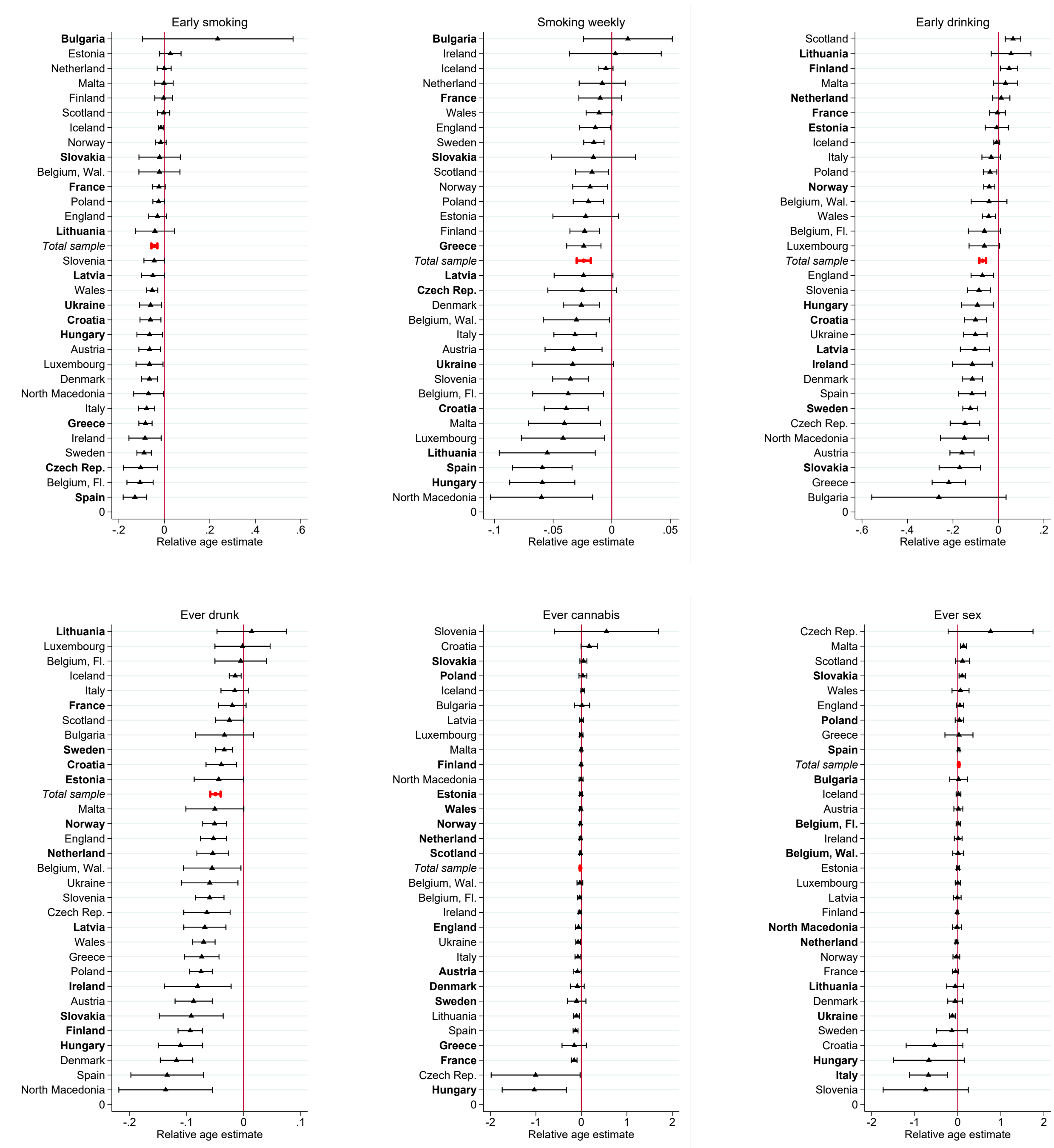
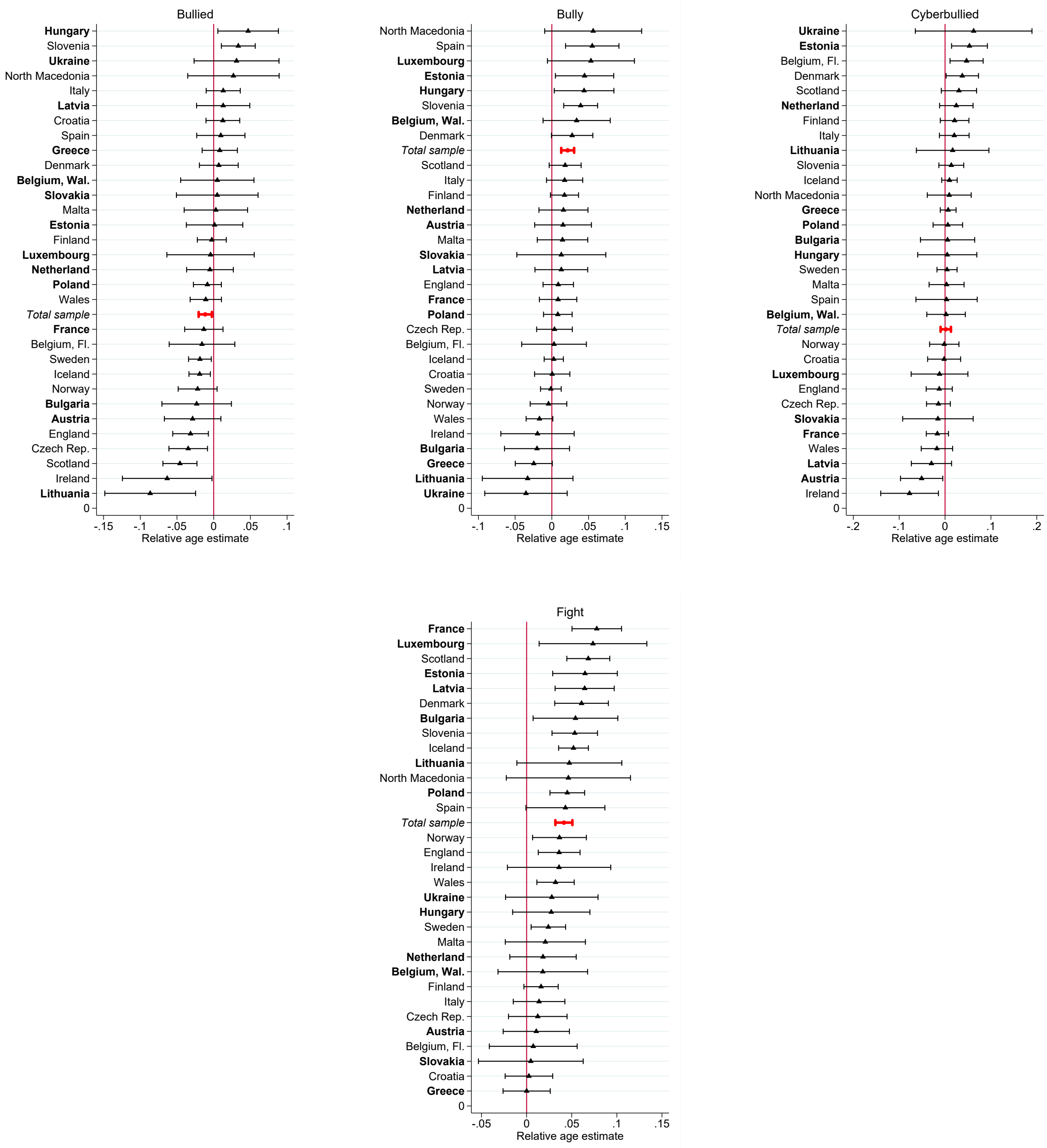


Figure A.3: Relative age effects on fighting behaviors.



B Appendix - Additional Tables

Table B.1: Cutoff dates, and quantity of observations per country per wave, for the HBSC dataset.

Country	Cutoff date	Wave 2001/2	2005/6	2009/10	2014/15	2017/18	All waves
Austria	Sep 1st	4,150	4,757	4,679	3,313	3,794	20,693
Belgium, Flanders	Jan 1st	1,345	3,113	3,029	3,230	3,013	13,730
Belgium, Wallonia	Jan 1st	3,026	3,589	3,080	4,845	4,146	18,686
Bulgaria	Jan 1st	0	4,826	0	4,639	3,199	12,664
Croatia	Apr 1st	4,270	4,680	6,058	5,507	4,687	25,202
Czech Republic	Sep 1st	5,006	0	4,324	5,041	11,265	25,636
Denmark	Jan 1st	4,468	5,319	3,921	3,784	3,112	20,604
England	Sep 1st	3,822	4,697	3,437	5,261	3,084	20,301
Estonia	Oct 1st	3,279	4,202	4,131	4,001	4,592	20,205
Finland	Jan 1st	5,143	5,143	6,496	5,810	0	22,592
France	Jan 1st	7,393	5,736	5,457	5,168	8,599	32,353
Greece	Jan 1st	0	0	4,808	4,078	3,807	12,693
Greenland	Jan 1st	0	0	198	141	556	895
Hungary	Jul 1st	3,985	3,450	4,569	3,737	3,456	19,197
Iceland	Jan 1st		8,480	8,747	9,160	3,643	30,030
Ireland	Jan 1st	1,951	3,730	1,859	3,366	3,120	14,026
Italy	Jan 1st	4,313	3,867	4,734	3,906	4,025	20,845
Latvia	Jan 1st	3,225	4,096	4,053	4,924	3,946	20,244
Lithuania	Jan 1st	5,586	5,575	5,221	0	1,507	17,889
Luxembourg	Sep 1st	0	2,889	2,968	2,192	2,315	10,364
Malta	Jan 1st	1,853	0	0	2,227	1,936	6,016
Netherlands	Oct 1st	3,778	3,796	4,076	3,862	4,206	19,718
North Macedonia	Jan 1st	3,593	4,749	3,434	4,096	4,072	19,944
Norway	Jan 1st	4,943		4,050	3,144	2,891	15,028
Poland	Jul 1st	6,245	5,475	4,190	4,068	4,953	24,931
Scotland	Mar 1st	4,381	6,130	6,668	5,672	4,799	27,650
Slovakia	Sep 1st	0	0	4,468	4,997	0	9,465
Slovenia	Jan 1st	3,894	5,070	5,322	4,795	5,574	24,655
Spain	Jan 1st	5,418	7,738	3,861	3,442	2,938	23,397
Sweden	Jan 1st	3,778	4,332	6,627	7,471	4,076	26,284
Ukraine	Jan 1st	3,943	4,859	5,345	3,095	5,263	22,505
Wales	Sep 1st	3,771	4,384	5,326	5,050	0	18,531
All countries		106,559	124,682	135,136	134,022	116,574	616,973

Note: Flanders and Wallonia as well as Denmark mainland and Greenland hold separate surveys within Belgium and Denmark, respectively.

Table B.2: First stage.

Variables	RA	AA
	(1)	(2)
ERA 1	0.001	-0.021***
	(0.004)	(0.003)
ERA 2	-0.040***	-0.062***
	(0.003)	(0.002)
ERA 3	-0.078***	0.037***
	(0.004)	(0.002)
ERA 4	-0.154***	-0.004**
	(0.003)	(0.002)
ERA 5	-0.205***	-0.040***
	(0.004)	(0.002)
ERA 6	-0.198***	0.043***
	(0.004)	(0.002)
ERA 7	-0.243***	0.011***
	(0.004)	(0.002)
ERA 8	-0.322***	-0.043***
	(0.005)	(0.002)
ERA 9	-0.346***	0.020***
	(0.005)	(0.003)
ERA 10	-0.336***	0.010***
	(0.005)	(0.003)
ERA 11	-0.341***	-0.004
	(0.006)	(0.003)
EAA	0.015***	0.990***
	(0.001)	(0.001)
N	577,691	596,387

Note: Clustered standard errors at the level of the class in parentheses. ERA stands for expected relative age. EAA stands for expected absolute age. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.3: RAE on smoking and drinking.

Variables	Smoking		Drinking	
	Early		Early	Ever drunk
	(1)	(2)	(3)	(4)
RA	-0.044***	-0.024***	-0.069***	-0.050***
	(0.008)	(0.004)	(0.009)	(0.005)
AA	0.053***	0.042***	0.042***	0.102***
	(0.001)	(0.000)	(0.001)	(0.001)
Female	-0.046***	-0.005***	-0.075***	-0.046***
	(0.002)	(0.001)	(0.002)	(0.002)
Parents	-0.080***	-0.047***	-0.060***	-0.076***
	(0.002)	(0.001)	(0.002)	(0.002)
SES: Medium	-0.006***	-0.009***	0.022***	0.006***
	(0.002)	(0.001)	(0.003)	(0.002)
SES: High	-0.005**	-0.009***	0.038***	0.021***
	(0.002)	(0.001)	(0.002)	(0.002)
N	285,742	464,318	285,190	448,722
Ancillary tests				
Underid. test: Lagrange	447.8	377.4	444.1	375.4
multiplier st. [p-value]	[0.001]	[0.001]	[0.001]	[0.001]
Weak id.: F-statistics	432.1	740.4	441.3	723.7
Overid. test: Hansen	17	7.780	17.51	5.590
J st. [p-value]	[0.074]	[0.650]	[0.064]	[0.848]

Note: Second stage estimates from the 2SLS. RA is relative age, AA is absolute age and it is centered around the mean to accommodate a more meaningful interpretation. SES is socio-economic status. Control variables include additionally include vectors for wave, country, and season-of-birth fixed effects. Clustered standard errors at the level of class in parentheses. Underid. stands for underidentification; Lagrange multiplier st. stands for Lagrange multiplier statistic; weak id. stands for weak identification; Overid. stands for overidentification; Hansen J st. stands for Hansen J statistics. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.4: RAE on marijuana consumption and sex.

Variables	Marjuana		Sex	
	Early	Ever	Ever	Unprotected
	(1)	(2)	(4)	(5)
RA	-0.014**	-0.022***	0.022**	0.055**
	(0.007)	(0.007)	(0.009)	(0.022)
AA	0.008***	0.060***	0.082***	-0.094***
	(0.001)	(0.001)	(0.002)	(0.009)
Female	-0.015***	-0.043***	-0.049***	0.060***
	(0.002)	(0.002)	(0.002)	(0.005)
Parents	-0.021***	-0.072***	-0.095***	0.027***
	(0.002)	(0.002)	(0.003)	(0.005)
SES: Medium	-0.001	0.008***	0.009***	-0.030***
	(0.002)	(0.002)	(0.003)	(0.006)
SES: High	0.001	0.017***	0.025***	-0.061***
	(0.002)	(0.002)	(0.002)	(0.006)
N	45,247	230,232	194,101	42,477
Ancillary tests				
Underid. test: Lagrange	194.7	344.6	387.3	327.9
multiplier st. [p-value]	[0.001]	[0.001]	[0.001]	[0.001]
Weak id.: F-statistics	160.9	308.2	370.5	181.9
Overid. test: Hansen	11.74	22.03	9.018	22.05
J st. [p-value]	[0.303]	[0.015]	[0.530]	[0.015]

Note: Second stage estimates from the 2SLS. RA is relative age, AA is absolute age and it is centered around the mean to accommodate a more meaningful interpretation. SES is socio-economic status. Control variables include additionally include vectors for wave, country, and season-of-birth fixed effects. Clustered standard errors at the level of the class in parentheses. Underid. stands for underidentification; Lagrange multiplier st. stands for Lagrange multiplier statistic; weak id. stands for weak identification; Overid. abbreviates overidentification; Hansen J st. stands for Hansen J statistics. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.5: RAE on aggressive behaviors.

Variables	Bullied (1)	Bully (2)	Cyberbullied (3)	Fight (4)
RA	-0.012**	0.022***	0.002	0.041***
	(0.006)	(0.005)	(0.006)	(0.006)
AA	-0.023***	0.009***	0.002***	-0.024***
	(0.000)	(0.000)	(0.001)	(0.001)
Female	-0.030***	-0.121***	0.029***	-0.300***
	(0.001)	(0.001)	(0.002)	(0.002)
Parents	-0.041***	-0.040***	-0.037***	-0.060***
	(0.001)	(0.001)	(0.002)	(0.002)
SES: Medium	-0.015***	0.004***	-0.002	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
SES: High	-0.023***	0.019***	-0.001	0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
N	566,168	566,196	224,131	561,632
Ancillary tests				
Underid. test: Lagrange	372.9	372.1	250.5	376
multiplier st. [p-value]	[0.001]	[0.001]	[0.001]	[0.001]
Weak id.: F-statistics	809.2	828	354	796.6
Overid. test: Hansen	20.47	13.54	5.008	4.373
J st. [p-value]	[0.025]	[0.195]	[0.891]	[0.929]

Note: Second stage estimates from the 2SLS. RA is relative age, AA is absolute age and it is centered around the mean to accommodate a more meaningful interpretation. SES is socio-economic status. Control variables include additionally include vectors for wave, country, and season-of-birth fixed effects. Clustered standard errors at the level of the class in parentheses. Underid. stands for underidentification; Lagrange multiplier st. stands for Lagrange multiplier statistic; weak id. stands for weak identification; Overid. stands for overidentification; Hansen J st. stands for Hansen J statistics. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.6: Mediation analysis on currently smoking.

	A's, Mediator effects (1)	% Indirect effect $[(A*B)/C]*100$ (2)	Ignorability of the mediator: Hansen J statistics (3)
Academic self-concept	-0.039*** (0.001)	33.3%	8.630 [0.567]
Well-being	-0.028*** (0.001)	12.2%	9.320 [0.502]
Body image	-0.014*** (0.001)	1.7%	9.759 [0.462]
Evening out	-0.055*** (0.001)	-0.9%	12.740 [0.239]
Students' support	-0.008*** (0.001)	3.5%	9.701 [0.467]
Main analyses, from Table 3			
RA, C	-0.024*** (0.004)		
Overidentification test: Hansen J statistics	7.780 [0.650]		

Note: Ignorability of the mediator with ignorability of the treatment (i.e., overidentification test from the main analyses) gives the sequential ignorability. All the analyses are obtained with the usual control variables: dummies for being female, for having both parents at home, for coming from medium or high SES family, as well as vectors for wave, country, season-of-birth fixed effects. In all the analyses the instruments are the eleven dummies for ERA and EAA. Clustered standard errors at the level of the class in parentheses. Hansen J statistics p-values in brackets. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.7: Mediation analysis on ever drunk.

	A's, Mediator effects (1)	% Indirect effect $[(A*B)/C]*100$ (2)	Ignorability of the mediator: Hansen J statistics (3)
Academic self-concept	-0.060***	25.7%	3.903
	(0.001)		[0.836]
Well-being	-0.034***	7.2%	4.368
	(0.001)		[0.839]
Body image	-0.030***	1.9%	4.655
	(0.002)		[0.806]
Evening out	-0.096***	-0.9%	4.794
	(0.001)		[0.844]
Students' support	-0.013***	2.8%	4.635
	(0.001)		[0.822]
Main analyses, from Table 3			
RA, C	-0.050***		
	(0.005)		
Overidentification test:	5.590		
Hansen J statistics	[0.848]		

Note: Ignorability of the mediator with ignorability of the treatment (i.e., overidentification test from the main analyses) gives the sequential ignorability. All the analyses are obtained with the usual control variables: dummies for being female, for having both parents at home, for coming from medium or high SES family, as well as vectors for wave, country, season-of-birth fixed effects. In all the analyses the instruments are the eleven dummies for ERA and EAA. Clustered standard errors at the level of the class in parentheses. Hansen J statistics p-values in brackets. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.8: Mediation analysis on early marijuana.

	A's, Mediator effects	% Indirect effect $[(A*B)/C]*100$	Ignorability of the mediator: Hansen J statistics
	(1)	(2)	(3)
Academic self-concept	-0.012***	12%	13.74
	(0.001)		[0.185]
Well-being	-0.009***	3.3%	13.63
	(0.001)		[0.190]
Body image	-0.008***	-0.2%	14.61
	(0.002)		[0.147]
Evening out	-0.020***	0.8%	14.97
	(0.001)		[0.133]
Students' support	-0.008***	6.2%	14.20
	(0.001)		[0.164]
Main analyses, from Table 3			
RA, C	-0.014***		
	(0.008)		
Overidentification test:	17.74		
Hansen J statistics	[0.303]		

Note: Ignorability of the mediator with ignorability of the treatment (i.e., overidentification test from the main analyses) gives the sequential ignorability. All the analyses are obtained with the usual control variables: dummies for being female, for having both parents at home, for coming from medium or high SES family, as well as vectors for wave, country, season-of-birth fixed effects. In all the analyses the instruments are the eleven dummies for ERA and EAA. Clustered standard errors at the level of the class in parentheses. Hansen J statistics p-values in brackets. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.9: Mediation analysis on ever sex.

	A's, Mediator effects (1)	% Indirect effect $[(A * B)/C] * 100$ (2)	Ignorability of the mediator: Hansen J statistics (3)
Academic self-concept	-0.052***	31.6%	10.13
	(0.001)		[0.429]
Well-being	-0.022***	-8.5%	10.21
	(0.002)		[0.422]
Body image	0.002	-0.1%	9.428
	(0.002)		[0.492]
Evening out	0.100***	-2.2%	9.008
	(0.001)		[0.531]
Students' support	-0.017***	-8.8%	9.155
	(0.001)		[0.517]
Main analyses, from Table 3			
RA, C	0.022***		
	(0.009)		
Overidentification test:	9.018		
Hansen J statistics	[0.530]		

Note: Ignorability of the mediator with ignorability of the treatment (i.e., overidentification test from the main analyses) gives the sequential ignorability. All the analyses are obtained with the usual control variables: dummies for being female, for having both parents at home, for coming from medium or high SES family, as well as vectors for wave, country, season-of-birth fixed effects. In all the analyses the instruments are the eleven dummies for ERA and EAA. Clustered standard errors at the level of the class in parentheses. Hansen J statistics p-values in brackets. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.10: Mediation analysis on fight.

	A's, Mediator effects	% Indirect effect $[(A*B)/C]*100$	Ignorability of the mediator: Hansen J statistics
	(1)	(2)	(3)
Academic	-0.058***	-28.6%	12.16
Self-concept	(0.001)		[0.274]
Well-being	-0.035***	-9.3%	13.10
	(0.001)		[0.218]
Body image	-0.043***	-3.1%	12.94
	(0.002)		[0.227]
Evening out	-0.072***	-1%	11.58
	(0.001)		[0.314]
Students' support	-0.036***	-9.3%	12.66
	(0.001)		[0.243]
Main analyses, from Table 3			
RA, C	0.041***		
	(0.006)		
Overidentification test:	6.373		
Hansen J statistics	[0.929]		

Note: Ignorability of the mediator with ignorability of the treatment (i.e., overidentification test from the main analyses) gives the sequential ignorability. All the analyses are obtained with the usual control variables: dummies for being female, for having both parents at home, for coming from medium or high SES family, as well as vectors for wave, country, season-of-birth fixed effects. In all the analyses the instruments are the eleven dummies for ERA and EAA. Clustered standard errors at the level of the class in parentheses. Hansen J statistics p-values in brackets. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table B.11: Cutoff dates, and quantity of observations per country per wave, for the ESPAD dataset, with indication about the additional presence in the HBSC dataset.

Country	Cutoff date	Wave 1995	Wave 2007	Wave 2011	Wave 2015	All waves	Also in HBSC
Albania	Jan 1st	0	0	3,165	2,519	5,684	
Austria	Sep 1st	0	2,559	0	3,656	6,215	x
Belgium, Flanders	Jan 1st	0	1,888	1,798	1,771	5,457	x
Bosnia and Herz., Fed. of Bosnia and Herz.	Apr 1st	0	2,973	4,489	0	7,462	
Bosnia and Herz., Serb Republic	Apr 1st	0	2,609	3,074	0	5,683	
Bulgaria	Jan 1st	0	2,323	0	0	2,323	x
Cyprus	Sep 1st	0	0	4,071	2,067	6,138	
Czechia	Sep 1st	0	3,901	3,864	2,710	10,475	x
Denmark	Jan 1st	2,209	867	0	0	3,076	x
Estonia	Oct 1st	0	2,314	2,402	0	4,716	x
Faroes Islands	Jan 1st	471	0	0	0	471	
Finland	Jan 1st	2,148	0	0	0	2,148	x
France	Jan 1st	0	2,880	2,529	2,598	8,007	x
Germany, Bavaria 1st	Oct	0	811	723	858	2,392	
Greece	Jan 1st	0	3,030	5,841	3,169	12,040	x
Ireland	Jan 1st	0	0	0	1,437	1,437	x
Italy	Jan 1st	1,409	0	0	0	1,409	x
Kosovo	Jan 1st	0	0	2,312	0	2,312	
Latvia	Jan 1st	0	2,236	2,592	0	4,828	x
Liechtenstein	Jul 1st	0	0	0	315	315	
Lithuania	Jan 1st	0	2,397	2,474	2,536	7,407	x
Moldova	Nov 1st	0	3,127	1,963	2,554	7,644	
Netherlands	Oct 1st	0	2,091	2,044	1,684	5,819	x
Norway	Jan 1st	3,807	0	0	0	3,807	x
Poland	Jul 1st	7,278	0	0	0	7,278	x
Russian Fed., Moscow 1st	Sep	0	1,963	1,718	0	3,681	
Slovakia	Sep 1st	2,287	2,443	1,949	2,169	8,848	x
Slovenia	Jan 1st	2,420	3,044	3,168	3,473	12,105	x
Sweden	Jan 1st	3,467	0	0	0	3,467	x
Ukraine	Jan 1st	6,365	2,411	2,210	2,259	13,245	x
All countries		31,861	45,867	52,386	35,775	165,889	

Note: Serb Republic is one of the two entities of Bosnia and Herzegovina, it is not a different name for Serbia.

Table B.12: Perceptions about risky behaviors.

Variable	N	Mean
Perceived risk of harming themselves		
Occasional cigarette	153,744	0.412
4-5 drinks a week	152,196	0.894
Occasional marijuana	144,503	0.756
Try ecstasy	133,925	0.751
Try amphetamine	135,651	0.944
Perceived peers' consumption prevalence		
Smoke	129,093	0.700
Drink	128,732	0.783
Drunk	124,021	0.478
Marijuana	128,113	0.147
Tranquilizers	126,078	0.033
Ecstasy	128,315	0.031
Inhaler	127,383	0.035

Note: All these outcomes on perceptions about risky behaviors are dummy variables.

Table B.13: Descriptive statistics for expected relative age, expected absolute age, control variables.

Variable	N	Mean	SD	Min	Max
Expected relative age	165,889	0.498	0.305	0	1
Expected absolute age	165,889	15.969	0.408	14.090	16.926
Female	165,889	0.514			
Parents	163,828	0.754			
SES: Low	131,868	0.099			
SES: Medium	131,868	0.512			
SES: High	131,868	0.389			

Note: SES: Low is the reference dummy for family SES. Female, Parents, and the SES variables are dichotomous variables; thus, minimum (min) and maximum (max) value, as well as standard deviation (SD), are not reported. Analyses additionally include vectors for wave, country and season of birth.