The Determinants of Sibling Similarity in Risky Behavior: Evidence from South Korea

Min-Jung Kim

Updated on November 12, 2025

Click here for the latest version

Abstract

Sibling similarity in adolescent risky behaviors is well established, yet its underlying sources remain unclear. This paper provides the first empirical evidence from South Korea that this similarity cannot be fully explained by parental socioeconomic status or direct sibling spillovers, and that unobserved common factors, such as parenting style and household norms, also play an essential role in shaping it. Using four complementary identification strategies, I find that socioeconomic background accounts for part of the observed correlation but can sometimes suppress the mechanisms that generate it. The influence of older siblings is modest, largely contemporaneous, and can discourage rather than imitate certain behaviors. A substantial portion instead reflects these unobserved common factors, notably, 67% for drinking and 86% for violence, which are more readily shaped by parental conduct and household norms. These effects are most pronounced among brother pairs and siblings close in age. This finding underscores that sibling similarity in risky behaviors is structured less by peer imitation than by family norms and intra-household dynamics.

Keywords Sibling correlation · Risky behavior · Family resource

Min-Jung Kim
E-mail: min24@hawaii.edu
Department of Economics, University of Hawai'i at Mānoa, Honolulu, HI, USA

1 Introduction

Adolescent risky behavior has long been recognized as a critical social issue across countries and over time.[1] Although often perceived as temporary deviations during adolescence, such behaviors can have lasting consequences for educational attainment, health outcomes, and subsequent labor market trajectories (Balsa et al., 2011;Abboud et al., 2024). Self-directed risky behavior, such as smoking and drinking, impairs academic achievement and raises the probability of school dropout (Cook and Moore, 1993), while antisocial risky behavior, including aggression and bullying, disrupts peer relationships and fosters social isolation (Carrell and Hoekstra, 2010).[2] These adverse dynamics accumulate over time, reducing educational opportunities (Black et al., 2013) and deteriorating health (Grossman, 2005). At the aggregate level, the persistence of risky behavior imposes substantial social costs by increasing welfare expenditures and heightening public spending on crime prevention (Gruber, 2009). Thus, risky behavior should not be understood merely as a transitory stage of adolescence, but as a key mechanism shaping individual life-course outcomes, the intergenerational transmission of inequality, and broader societal costs. This underscores the importance of carefully identifying its determinants and understanding the channels through which such behaviors persist and propagate.

Attempts to analyze the determinants of risky behavior at the individual level soon encounter fundamental limitations. Even among adolescents who share similar socioeconomic backgrounds and attend the same schools within the same neighborhoods, the manifestation of risky behavior diverges substantially, and striking discrepancies are frequently observed even within the same household. This implies the existence of intra-household dynamics that cannot be fully accounted for by individual traits or external conditions alone. To capture these dynamics, it is particularly suitable to expand the unit of analysis to sibling pairs. Siblings are simultaneously exposed to the shared environment, such as parental socioeconomic status, parenting style, and household norms, while also experiencing non-shared environments, including differential treatment by birth order, gender, or age gap. For instance, when parents work long hours, both siblings may face limited supervision, but the timing and intensity of risky behavior initiation can differ depending on gender or age. Examining sibling similarity in risky behavior thus provides not merely correlational evidence, but also a methodological strategy to identify structural family-level factors and intra-household transmission mechanisms. Consistent with this perspective, prior studies have documented strong sibling correlations across various dimensions, including smoking (Harris and López-Valcárcel, 2008), drinking (Koch and Ribar, 2001), educational attainment and school absence (Nicoletti and Rabe, 2019), as well

[1] According to the World Health Organization (WHO), approximately 10% of adolescents aged 13-15 worldwide report tobacco use, and more than one-quarter of those aged 15-19 are classified as current drinkers. Moreover, interpersonal violence during adolescence is identified as one of the leading causes of mortality, closely linked to physical injury, deteriorating mental health, school dropout, and sexual and reproductive health problems.(World Health Organization (2024)

[2] In the Korean context, Lee and Hoe (2018) finds that adolescents with risky behavior experience report higher levels of aggression and social withdrawal compared to their peers, and that this social withdrawal, reflecting weakened peer relationships and social isolation, is structurally linked to lower school adjustment.

as sexual and aggressive risky behavior (Rodgers and Rowe, 1988;Tippett and Wolke, 2015).

Despite extensive evidence of sibling similarity across various socioeconomic outcomes, the underlying mechanisms that generate such similarity remain difficult to disentangle. Simple correlations cannot reveal whether similarity arises from (i) shared parental socioeconomic status (SES), which jointly determines siblings' access to resources and opportunities, (ii) direct behavioral spillovers, whereby the older child's actions influence the younger through imitation, competition, or deterrence, or (iii) common environmental factors, such as parenting style or household norms, that induce similar responses in both children. Without explicitly separating these channels, sibling similarity remains a reduced-form statistic rather than evidence of a specific mechanism. Existing research has acknowledged the multiplicity of explanations. Some studies emphasize the role of parental SES as a common determinant of opportunity structures, arguing that resemblance primarily reflects parallel resource allocation within families (Conley and Glauber, 2005). Others highlight the role of older siblings as social role models or rivals, showing that interventions targeting one child may propagate to others via intra-household externalities (Booth and Kee, 2009). A third line of work suggests that parents adjust expectations and disciplinary strategies after observing the first child's behavior, implying that similarity may stem from adaptive parental beliefs rather than direct imitation(Mechoulan and Wolff, 2015). While existing studies point to multiple plausible channels through which sibling similarity may arise, their relative importance has not been empirically separated. This identification gap motivates this study.

This paper investigates whether sibling similarity in risky behavior reflects mere correlation or distinct structural mechanisms. While parental socioeconomic status (SES), behavioral spillovers between siblings, and unobserved family influences may all generate similarity, these forces differ fundamentally in their temporal structure and interaction patterns. They therefore cannot be treated as a single mechanism. The parental SES represents a simultaneous and nondirectional exposure shared by both siblings, whereas spillovers imply a sequential and asymmetric transmission based on birth order. Latent family norms, in contrast, reflect persistent and jointly enforced patterns that operate beneath observable behavior. Treating these distinct forces as a single mechanism obscures their individual contributions; therefore, they must be explicitly disentangled. Accordingly, this study addresses the following three questions. First, to what extent does SES contribute to sibling similarity in risky behavior? Second, does the observed resemblance reflect genuine temporal diffusion of behavior across siblings? Third, is it instead driven by unobserved family common factors? In addition, I examine how these mechanisms vary with sibling gender composition and age spacing, using data on Korean adolescents.

To address these questions, I draw on the Korean Children and Youth Panel Survey 2018 (KCYPS 2018), the first nationally representative panel dataset in Korea that simultaneously collects information on 2005-born adolescents and their siblings. This dataset provides a unique opportunity to empirically identify the structural sources of sibling similarity in risky behavior. In particular, the KCYPS 2018 includes rich self-reported measures of multiple risky behaviors as well as detailed household background information, making it especially well-suited for analyzing sibling dynamics in the Korean context, where educational competition is intense and parental investments and norms are strongly differentiated by birth order and gender. [3]

This study adopts a plausible directional structure in which behavior is assumed to flow from older to younger siblings.[4] This assumption is not presented as proof of causality but rather as a theoretically grounded premise for empirical identification. It is motivated by consistent empirical evidence showing that older siblings tend to initiate risky behaviors earlier and engage in them at higher rates. This directional framework aligns with established findings on birth-order effects and behavioral transmission through imitation or competition, and thus provides a reasonable starting point for empirical identification.[5] Accordingly, the analysis evaluates the full sample and heterogeneity across sibling gender composition and age gap to assess where such transmission patterns appear stronger or weaker.

As a first step, I apply a mixed-effects model to quantify how much sibling similarity in risky behavior can be attributed to parental socioeconomic status (SES). Because parental SES affects both siblings at the same time and in similar ways, it represents a shared background factor rather than a directional influence. Therefore, it cannot be interpreted as a causal pathway in the same way as a spillover. Accordingly, its contribution is most appropriately inferred from the covariance structure. This approach, widely used in the literature, allows me to assess the contribution of the shared environment across siblings.[6] The results show that parental SES accounts for only a limited share of similarity across the full sample as well as across gender composition and age gap. In some cases, suppressor effects or overestimation of SES contributions are observed. This suggests that sibling similarity cannot be fully explained by resource allocation alone and that additional structural forces must be at play.

To address the reflection problem (Manski, 1993)[7], which refers to the empirical difficulty of distinguishing true behavioral spillovers from shared family influences, I estimate a joint dynamic probit model following Altonji et al. (2017) to identify

[3] Korean Children and Youth Panel Survey 2018 (KCYPS 2018) includes self-reported measures of smoking, drinking, truancy, sexual behavior, and violence, and its longitudinal design makes it possible to analyze how the past behaviors of the older sibling affect the subsequent behaviors of the younger sibling. Because data constraints have severely limited research on sibling dynamics among Korean adolescents, this study is the first to leverage such data to identify the structural determinants of sibling similarity in risky behaviors within the Korean context.

[4] As shown in A1, older siblings consistently exhibit higher rates of each risky behavior than younger siblings within the dataset.

[5] for example, Argys et al. (2006) examines birth order and adolescent risky behavior, and Averett et al. (2011) also analyze the role of parenting in mediating the effect of older siblings on younger siblings' risky behavior.

[6] Using this framework,Mazumder (2008) finds that parental income and education explain little of sibling similarity in addictive behaviors, suggesting that socioeconomic status plays only a limited role in shaping delinquent or addictive behaviors. Similarly, Eriksson et al. (2016) shows that parental criminal history, rather than income or education, accounts for a larger share of correlations in criminal behavior among siblings.

[7] Manski (1993) distinguishes three sources of confounding in peer effect models: (i) contextual effects, which arise when individuals respond to the characteristics of their peers, (ii) correlated effects, which occur when individuals in the same group behave similarly due to shared environments or unobserved common factors, and (iii) endogenous effects, which capture the simultaneous influence of peers on each other's behavior.

whether behavioral transmission truly occurs within sibling pairs. Because genuine spillover requires sequential timing or role asymmetry, this framework enables a dynamic evaluation of whether the older sibling's behavior affects the younger sibling's subsequent choices. The estimates indicate that resemblance arises primarily contemporaneously, while lagged effects are limited. Although some heterogeneity appears across gender composition and age gap, the overall magnitude is modest, suggesting that the direct influence of older siblings may be relatively small. In other words, sibling similarity cannot be attributed solely to imitation or sequential contagion, but is likely to be mediated by unobserved family norms or parenting styles.

Building on this, I adopt a correlated random effects (CRE) framework, following the approach of Altonji et al. (2017)[8], to estimate the role of unobserved common factors such as household norms and parenting styles. The results show that risky behaviors that are more easily observable to parents, such as drinking and violence, are largely explained by unobserved common factors. Once these unobserved common factors are accounted for, the direct effect of older siblings becomes limited and, in some cases, even resembles a suppressing effect. These findings confirm that the relative importance of unobserved family dynamics varies across different types of risky behavior, with parenting style and household environment playing a central role in shaping sibling similarity.

In the final step, I implement a decomposition model to disentangle the quantitative contributions of three channels, which consist of parental SES, older sibling spillovers, and unobserved common factors. This approach is designed to separate the pure effects of each channel from its interactions, thereby synthesizing the qualitative insights from the preceding models into a unified quantitative framework. The results consistently indicate that parental SES and unobserved factors primarily drive sibling similarity in risky behavior. In contrast, the spillover effect of older siblings is relatively limited. In particular, for behaviors such as drinking and violence, where parental norms and household practices are more directly operative, the contribution of common unobserved factors is especially pronounced. Moreover, a comparison between the fixed effects (FE) and the correlated random effects (CRE) specifications reveals that the explanatory power of SES tends to be overstated in FE models. These findings underscore the insufficiency of explaining sibling correlations in risky behavior solely through observable factors.

Taken together, the findings show that while parental socioeconomic status has some explanatory power for sibling similarity in risky behavior, its contribution is not consistent. Instead, unobserved common factors appear to play a central role in shaping sibling resemblance. Moreover, the transmission from older to younger siblings emerges only in limited and conditional ways depending on the visibility of the behavior, sibling gender composition, and age spacing. These results suggest that sibling similarity in risky behavior cannot be reduced to either household economic background or simple imitation. Rather, they underscore the structural importance of unobserved intra-family dynamics, such as parenting style and household norms.

In this context, the implications of this study extend beyond simply documenting

[8] Altonji et al. (2017) show that, in their model, much of the observed similarity in sibling substance use is explained not by the pure influence of the older sibling but rather by unobserved family-specific factors.

the existence of sibling correlations. The observed structure of similarity not only raises the question of "who influences whom," but also provides a deeper interpretation of whether parental practices within the household mediate the formation of children's behaviors. In particular, the finding that older sibling spillover effects are limited while the contribution of common family factors remains consistently large suggests that sibling interactions cannot be understood through a single mechanism. The same family environment does not operate uniformly across all children. Rather, transmission is mediated by sibling gender composition, age spacing, and the specific nature of the behavior. This reinforces the notion that adolescent risky behavior should therefore be analyzed at the household level, acknowledging the heterogeneous structure within families.

Accordingly, this paper makes two distinct contributions. First, it offers a substantive contribution by providing the first structural analysis of sibling similarity in adolescent risky behavior within the South Korean context. Prior research in South Korea has largely followed the international literature by focusing on external determinants such as parental socioeconomic status (Aizer, 2004), neighborhood conditions (Kling et al., 2005), or peer influence (Lundborg, 2006), while largely excluding within-family role structures and transmission mechanisms from the analytical framework. Even the few studies that relied on sibling samples typically summarized resemblance using a single correlation or an average spillover coefficient, without decomposing its internal structure. This paper departs from that approach by redefining sibling resemblance not as a reduced-form statistic but as the composite outcome of multiple structural mechanisms.

Second, this research expands empirical identification in the peer-effects literature by proposing a sequential decomposition strategy that distinguishes behavioral transmission from similarity. Existing studies of sibling or peer influence face a fundamental reflection problem, which is that observed similarity is inherently endogenous, as it is difficult to separate causal spillover from shared exposure or correlated responses. As a result, prior studies have often relied on strong symmetry assumptions or shock-based instruments that capture only local variation, and spillover effects are prone to upward bias when shared environments or unobserved family norms are not properly netted out. This paper addresses this limitation by combining a mixed-effects model to partial out covariance induced by socioeconomic background, a joint dynamic probit specification to track temporal diffusion, and a correlated random effects framework to quantify latent within-family norms. Rather than depending on a single exclusion restriction, the identification proceeds by sequentially stripping away SES, spillover, and unobserved common factors, offering a generalizable framework for disentangling transmission from shared exposure in intra-household and peer settings.

In contemporary South Korean society, where family structures and parenting practices are undergoing rapid transformation, adolescent risky behavior remains a pressing social concern, closely linked to adverse outcomes such as persistently high suicide rates.[9] South Korean adolescents are uniquely exposed to strong familial ex-

[9] According to the study based on the 12th (2016) Korea Youth Risk Behavior Web-based Survey, cigarette smoking and smoking intensity are significant predictors of suicidal ideation, suicide planning, and suicide attempts among Korean adolescents.(Kim et al. (2017)

pectations and norms, which distinguishes their experience from that of youth in other countries. By situating the analysis in this context, the study provides a more precise account of the structural factors shaping adolescent behavior and offers an analytical framework with potential relevance for mitigating youth delinquency and suicide risks.

The remainder of the paper is organized as follows. Section 2 describes the KCYPS 2018 data, the construction of the sibling sample, and descriptive patterns in risky behaviors. Section 3 outlines the empirical strategy and main findings: Section 3.1 estimates sibling correlations using a REML-mixed model to assess how parental socioeconomic resources shape similarity in risky behavior. Section 3.2 identifies causal peer effects from older to younger siblings with a joint dynamic probit model, distinguishing contemporaneous and lagged influences. Section 3.3 investigates the impact of shared family environments using a correlated random effects specification. Section 3.4 shows a structural decomposition to quantify the relative contributions of socioeconomic status, sibling spillovers, and unobserved common factors. Across all analyses, I additionally explore heterogeneity by sibling gender composition and age gap. Section 4 concludes by summarizing the results and discussing their policy and academic implications.

2 Data and Descriptive Statistics

2.1 Data

This study utilizes data from the Korean Children and Youth Panel Survey 2018 (KCYPS 2018), the panel dataset that records South Korean adolescents' growth and developmental processes. KCYPS 2018 began 2018 with its first wave, targeting first-year middle school students born in 2005. The data was administered annually for six waves, continuing through 2023, when the original cohort entered their final year of high school. Accordingly, this study employs all publicly available waves, from the 1st wave in 2018 to the 6th wave in 2023. The survey was conducted with three main groups, including primary respondents, students born in 2005, parents, and siblings. This is the only Korean panel dataset that systematically includes target respondents and their siblings, enabling me to link individual behavior to broader household context. In addition to rich information on the focal youth, the dataset offers valuable behavioral measures on siblings, allowing for the identification of intra-household dynamics and socioeconomic background factors.

Using household identifiers to construct the analytic sample, I match each focal youth to the parents and siblings identified within their household identity number. For consistency, the analysis is restricted to cases where the primary sample functions as the older sibling. Specifically, focal youth born in 2005 are classified as older siblings, and only their siblings born after 2005 are treated as younger siblings. This classification is based on two considerations. First, the primary sample is larger than the sibling sample, and most siblings are later-born children; thus, defining the analysis unit based on the primary sample ensures greater statistical stability and comparability. Second, defining birth order using the sibling dataset would create inconsistencies, as many siblings are born after the focal youth, and some older siblings are not fully observed in the data. Using the primary sample as the reference avoids these issues and allows for a clearly defined hierarchical sibling structure. In the case of twins, the birth order information is used to distinguish between them.

For analytical clarity and comparability, the sample is restricted to two children per household. This restriction reflects two underlying considerations. First, in households with more than one younger sibling, only the sibling closest in age to the primary sample is included. Younger siblings with a larger age gap are less likely to have reached a developmental stage at which risky behaviors can be observed, increasing the likelihood of missing values in outcome variables. Second, by imposing a minimum age threshold for younger siblings, the analysis secures observable behavioral data and enables comparisons within a comparable developmental range. Accordingly, only younger siblings who are at least in the first grade of elementary school are included. As a result, the analytic sample consists of exactly two children per household, enhancing the analysis's reliability and consistency.

The outcome variables used in this study are based on a broad set of self-reported problem behaviors, including smoking, drinking, school absence, running away, severe mocking, bullying, group fighting, physical violence, threatening, extortion, theft, sexual intercourse, sexual harassment, gambling, and verbal abuse. Based on existing literature, I focus on five widely studied and prevalent behaviors among Korean youth, which are smoking, drinking, school absence, sexual intercourse, and violence. [10] In terms of violence, which captures interpersonal behaviors that are physically or psychologically harmful to others. This includes severe mocking, bullying, group fighting, hitting, threatening, and verbal abuse.

In the original dataset, the response categories for the risky behavior variables are (1) Never, (2) 1-2 times per year, (3) Once a month, (4) 2-3 times a month, (5) Once a week, and (6) Multiple times a week. To focus on behavioral onset rather than intensity, I recode each variable into a binary indicator 1 if the respondent has engaged in the behavior at least once (score 2), and 0 if never. This extensive margin coding strategy emphasizes whether the behavior was ever initiated, which is consistent with an interesting point in this research on behavioral transmission rather than frequency. The existing studies have also adopted this approach and shown that the mere presence or absence of risky behavior is more relevant for analyzing sibling similarity and spillovers (Mazumder, 2008;Fletcher et al., 2012). For the Violence variable, I take the maximum value across its six constituent items to capture the most severe violent behavior observed. This maximum is converted into a binary variable following the same procedure.

Demographic information includes birth order, gender, age, number of siblings, residential areas, and parental socioeconomic characteristics such as education, age, employment status, and household income. Also, I consider the possibility that sibling similarity may arise not only from direct behavioral spillovers but also from unobserved shared characteristics, such as personality traits or parenting style, that differ across adolescents but are not directly measured. To mitigate this concern, I

[10] Several studies have identified smoking, drinking, school absence, sexual intercourse, and violent behaviors as key indicators for youth risky behavior factors in South Korea.(Kim et al., 2016;Valencia et al., 2019)

follow Altonji et al. (2017) and additionally control for individual-level factors that may influence risky behavior, including the adolescent's relationship with parents and peers, academic satisfaction, and overall school experience.

2.2 Descriptive Statistics

Table 1 presents household characteristics by sibling gender composition. According to Eriksson et al. (2016), the size and direction of sibling influence can vary depending on household structure, particularly sibling gender composition and age difference. Accordingly, this study approaches this classification to examine how heterogeneity in sibling composition is associated with similarity in risky behaviors. In the sample, the average age gap between siblings is 2.77 years. Based on this, the sample is divided into two groups: those with an age difference of two years or less and those with a gap of three years or more. Regarding gender composition, the sample consists of 1,959 households with two brothers, 1,594 with an older sister and a younger brother, 1,807 with two sisters, and 1,941 with an older brother and a younger sister. The average number of children per household in the full sample is 2.35, indicating that restricting the analysis to households with exactly two children does not compromise representativeness. Accordingly, the final analytic sample consists of 7,301 households, each including a focal youth born in 2005 and one younger sibling.

Parental education level, employment status, and household income are evenly distributed across sibling gender groups, suggesting little risk of systematic selection bias into gender composition based on observable socioeconomic factors. For instance, the share of fathers with a university degree ranges from 32% to 38%, and maternal labor force participation is consistently around 70%. The average monthly household income across groups ranges from approximately USD $4,263 to USD $4,803, confirming the balance of observable characteristics. These patterns support two possible interpretations. First, risky behavior is unlikely to be determined solely by observable parental socioeconomic characteristics associated with sibling gender composition. Second, unobserved parental inputs, such as differences in parenting styles, expectations, or behavioral norms, may systematically vary by the gender makeup of the children. Even among parents with similar education and income levels, those raising two sons may exhibit different parenting behavior than those raising two daughters or mixed-gender siblings. This implies that observed sibling similarity in risky behavior may stem from direct sibling spillovers and shared unobserved parental inputs.

Figure 1 summarizes the mean incidence of each risky behavior by gender, from which two salient patterns emerge. First, the prevalence of risky behaviors varies depending on their characteristics. The relative ranking of behaviors is similar for both boys and girls. Sexual intercourse is the least common behavior, with only 3% of boys and 2% of girls reporting any experience. Violence, on the other hand, is the most prevalent, reported by 18% of boys and 10% of girls. These differences are closely related to the visibility and transmission structure of each behavior. Sexual intercourse is highly private and individualized, which limits opportunities for direct transmission or observation. In contrast, violence tends to be more observable and

Table 1 Household Characteristics by Sibling Gender Composition

	All	Brother-only	Mixed (S&B)	Sister-only	Mixed (B&S)
Father's Background					
Education					
High School Graduate	0.19	0.19	0.17	0.19	0.19
	[0.39]	[0.40]	[0.38]	[0.40]	[0.39]
College Graduate	0.16	0.14	0.14	0.19	0.16
	[0.36]	[0.35]	[0.35]	[0.39]	[0.37]
University Graduate	0.33	0.32	0.38	0.33	0.3
	[0.47]	[0.47]	[0.49]	[0.47]	[0.46]
Working (Work =1)	0.92	0.92	0.94	0.9	0.91
	[0.27]	[0.27]	[0.24]	[0.29]	[0.28]
Mother's Background					
Education					
High School Graduate	0.23	0.23	0.23	0.24	0.22
	[0.42]	[0.42]	[0.42]	[0.43]	[0.41]
College Graduate	0.19	0.2	0.17	0.21	0.19
	[0.40]	[0.40]	[0.38]	[0.41]	[0.40]
University Graduate	0.26	0.23	0.29	0.26	0.25
	[0.44]	[0.42]	[0.46]	[0.44]	[0.43]
Working (Work =1)	0.72	0.73	0.74	0.68	0.72
	[0.45]	[0.44]	[0.44]	[0.47]	[0.45]
Household's Background					
Age of Older Sibling	15.5	15.51	15.48	15.5	15.51
	[1.71]	[1.71]	[1.71]	[1.71]	[1.71]
Age of Younger Sibling	12.73	12.75	12.75	12.72	12.7
	[2.20]	[2.21]	[2.22]	[2.17]	[2.21]
Age Gap	2.77	2.76	2.74	2.78	2.81
	[1.43]	[1.44]	[1.44]	[1.37]	[1.45]
Number of Siblings	2.35	2.36	2.32	2.43	2.29
	[0.59]	[0.62]	[0.61]	[0.58]	[0.55]
Household Income(US$)	5,000	4,741	4,661	4,803	4,263
	[2.11]	[2.08]	[2.03]	[2.20]	[2.10]
N	7,301	1,959	1,594	1,807	1,941

Notes. In South Korea, universities provide a full range of academic degrees and programs in 4 years, while colleges focus on shorter, specialized training and vocational education in 2 years. Household income is coded into 12 categorical brackets ranging from 0 to over 10 million KRW per month. Each income bracket is converted to a continuous value by assigning the interval's midpoint in Korean Won. For example, the seventh bracket is treated as 5.5 million KRW. The midpoint value is multiplied by 1,000,000 KRW and converted into USD using the annual average KRW/USD exchange rate for each survey year from 2018 to 2023. The exchange rates are based on yearly averages reported by Federal Reserve Economic Data of Louis (2024).

often occurs in interactive settings, making it more likely to be transmitted between siblings through observation, imitation, or indirect learning. Second, boys exhibit consistently higher rates of risky behavior across all categories. In particular, the incidence of smoking, drinking, and violence among boys is roughly twice that of girls. This suggests that boys are more likely to be exposed to or engage in risky behaviors. These gender differences might be especially pronounced in households with two sons, where intra-sibling dynamics may foster more intense behavioral interactions. The implication is that sibling spillovers and behavioral transmission mechanisms may differ significantly depending on gender composition.[11]

[11] For example, Dahl et al. (2024) found that same gender siblings tend to imitate each other's choices. In addition, Black et al. (2021) document that sibling spillovers can be shaped by gender composition and age proximity, suggesting that behavioral interactions may be stronger among brothers.

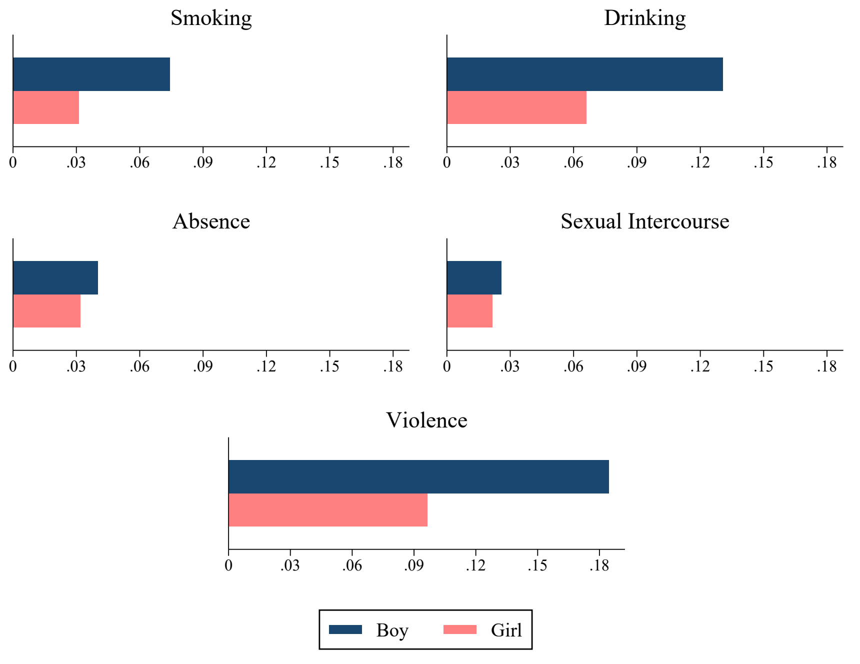

Fig. 1 Gender Differences in Average Risky Behavior
Notes. This figure shows the mean prevalence of each risky behavior by gender. The statistics are based on the full sample and do not condition on sibling birth order.

These descriptive insights underscore that, beyond risky behavior, examining sibling spillovers more broadly requires treating sibling gender composition and age differences as central analytical dimensions. Moreover, family structure is a structural context that shapes adolescents' behavioral development and socialization pathways, providing an essential lens for understanding intra-household behavioral mechanisms.

3 Empirical Strategies and Results

In this section, I conduct a four-stage empirical analysis to disentangle the mechanisms that generate sibling similarity in risky behaviors. First, I compare sibling correlations before and after controlling for parental socioeconomic status (SES) to assess the extent to which parental SES explains behavioral similarity among siblings. Second, I examine whether these correlations reflect not merely contemporaneous covariance but causal transmission, by testing whether the older sibling's past and current behaviors influence the younger sibling's current behavior. Third, I analyze the remaining correlation after accounting for these channels to identify the influence of unobserved common family factors such as household norms and parenting styles.

Finally, I decompose the overall correlation into these components to quantify the relative contribution of each mechanism to different types of risky behaviors.

This stepwise empirical strategy moves beyond comparing the magnitude of correlations and instead structurally decomposes behavioral similarity into three components, parental socioeconomic status (SES), the causal influence of the older sibling, and unobserved family factors, to identify the underlying causal mechanisms. However, these mechanisms are unlikely to operate uniformly across households. Prior research shows that gender composition shapes both the strength and direction of sibling interactions, with same-gender siblings exhibiting stronger imitation effects than mixed-gender pairs (Cools and Patacchini, 2017). Age gap likewise determines opportunities for observation and mutual learning, as spillover effects tend to be stronger when siblings are closer in age (Dahl et al., 2024). To capture this heterogeneity, I estimate all four stages of the analysis separately for brother–brother, sister–sister, and mixed-gender pairs, as well as for families with small (two years or less) and large (more than two years) age gaps.

By incorporating variation in family structure, this approach strengthens identification of the underlying mechanisms generating sibling resemblance. It allows for a clearer separation between behavioral similarity driven by shared socioeconomic environment, direct sibling spillovers, and unobserved family factors, and reveals how these channels differ across family contexts.

3.1 Parental Socioeconomic Resources and Sibling Correlations

3.1.1 Empirical Strategy: REML-Mixed Model

Diverse factors influence children's behavioral outcomes, but one of the most important explanatory variables is household resources, particularly parents' SES. The existing studies show that parents' SES has persistent effects on a variety of outcomes, including children's health (Case et al., 2002; Currie and Stabile, 2003), cognitive and emotional development (Currie, 2009), and educational attainment (Aizer et al., 2016). From this perspective, it is natural to hypothesize that long-term environmental conditions shared within families may be a key determinant of similarity in sibling behavior. To empirically examine how parents' SES can explain sibling similarity, I employ a mixed model following the methodological approaches of Solon et al. (1991) and Mazumder (2008). First, I estimate correlations in risky behavior between siblings and then analyze the extent to which parental education and income contribute to sibling correlations. The linear model used for this purpose can be expressed as follows:

$$R_{sft} = \beta \mathbf{X}_{sft} + \varepsilon_{sft} \tag{1}$$

where R_{sft} represents risky behavior for a sibling s in family f at year t. $\mathbf{X}_{sft}$ Vector of baseline observable characteristics for each sibling and error term.

$$\varepsilon_{sft} = \alpha_s + \mu_{sf} + \nu_{sft} \tag{2}$$

The error term consists of a time-invariant family component, a time-invariant sibling-specific component, and an idiosyncratic time-varying component. These three components of the error term are assumed to be independent, meaning that these components are assumed to be orthogonal to each other, implying that cross-level covariances are zero. Under this assumption, the variance of the error term can be expressed as follows:

$$var(\varepsilon_{sft}) = var(\alpha_s) + var(\mu_{sf}) + var(\nu_{sft}) = \sigma_\alpha^2 + \sigma_\mu^2 + \sigma_\nu^2 \tag{3}$$

This study focuses on time-invariant factors. Accordingly, the correlation coefficient of risky behavior between siblings is defined in Equation (4) as follows:

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\mu^2} \tag{4}$$

The sibling correlation coefficient, ρ, is the proportion of total variance in risky behavior attributable to time-invariant household-level factors. By construction, $\rho \leq 1$. A higher value of ρ indicates that time-invariant household resources influence sibling similarity more than transitory or individual-specific factors. Therefore, this analysis focuses on the role of stable, long-term family characteristics in shaping sibling behavior.

The measures of parents' SES used in this study are constructed as follows. Parental education includes both fathers and mothers' schooling, measured on a seven-point scale ranging from 1, which indicates no formal education, to 7, which indicates graduate school or higher. The intermediate categories include primary school, middle school, high school, two-year college, and four-year university completion. Parental income is reported on a twelve-point categorical scale, where 1 indicates no income and 12 corresponds to a monthly income above 10 million KRW. Each intermediate category, 2 through 11, reflects a one-million-KRW range. For instance, 1 to 2 million KRW, 2 to 3 million KRW, 3 to 4 million KRW, and so forth, thereby capturing differences in household financial resources.

Parental education and income are typically highly correlated. If only one of these variables is included in the model, the contribution of each factor may be over- or underestimated due to omitted variable bias. Accordingly, this study simultaneously controls for both education and income, thereby accounting for their correlation and mitigating biases and multicollinearity in the coefficient estimates. While including only education or only income does not dramatically alter the variance of the estimates, including both variables together yields a greater change in variance. This reflects a structural feature of the Korean context. According to summary statistics, approximately 68% of fathers have completed at least high school, indicating relatively compressed educational attainment. Thus, the education variable exhibits relatively low variance, whereas income demonstrates greater heterogeneity and dispersion. Including both variables, therefore, allows for a more refined decomposition of the sources of sibling similarity and provides more stable estimates of each factor's relative contribution.

3.1.2 Baseline Sibling Correlations

Figure 2 compares the estimated sibling correlations before and after controlling for parents' SES in the full sample, illustrating how parents' SES affects the magnitude of sibling similarity in risky behavior. The comparison reveals different patterns across behaviors. Smoking declines from 0.018 to – 0.004 and becomes statistically insignificant. Drinking decreases from 0.024 to 0.016 but remains significant. Violence falls slightly from 0.031 to 0.027. By contrast, school absence rises from 0.038 to 0.048, and sexual intercourse increases from 0.084 to 0.091 after controlling for parents' SES. The explained component of the change attributable to observed parents' SES is 122% for smoking, 31% for drinking, 25% for absence, –8% for sexual behavior, and 13% for violence.

Decomposing overall sibling correlations into the component explained by observed parents' SES suggests two key insights. First, in the case of smoking, the fact that the correlation change exceeds 100%, the coefficient reverses sign, and the non-significance indicates that parents' SES may appear to contribute excessively to sibling similarity. This points to the possibility of omitted variable bias. For example, if important determinants such as parents' smoking habits or attitudes toward children's health are excluded from the model, the coefficients on parental resources may be biased. In addition, since variables such as income and education are often self-reported, measurement error could have increased the instability of the estimates. Second, the increase in sibling similarity for sexual intercourse and absence after controlling for parents' SES suggests that parental resources may have had a suppressor effect. Rather than directly influencing the outcome variables, parents' SES may have mediated or moderated the influence of unobserved family-level factors negatively correlated with sibling similarity. For example, in the case of sexual intercourse, higher parental education or income could reduce sibling similarity if parents adopt differentiated parenting or monitoring strategies for each child or apply different values regarding sexuality. Similarly, in the case of absence, factors such as parental work patterns, parenting styles, or child supervision strategies may lead siblings to experience different outcomes, thereby contaminating estimates before and after controlling for parents' SES, reflecting the unobserved effect of parental rearing styles.

Overall, the results indicate that observable parental SES variables such as parental education and income might not substantially explain sibling similarity in risky behavior. Even after controlling for these observable characteristics, changes in sibling correlations remain small or inconsistent in direction, suggesting that socioeconomic resources alone are unlikely to be the primary determinants of similarity. Instead, the findings point to the importance of unobserved family processes, including parenting practices, household norms, and sibling interactions, that operate beneath measurable socioeconomic differences. In other words, visible parental economic and educational resources capture only part of the story, while the unobserved relational and behavioral dynamics within families appear to play a more fundamental role in shaping sibling resemblance in risky behaviors.

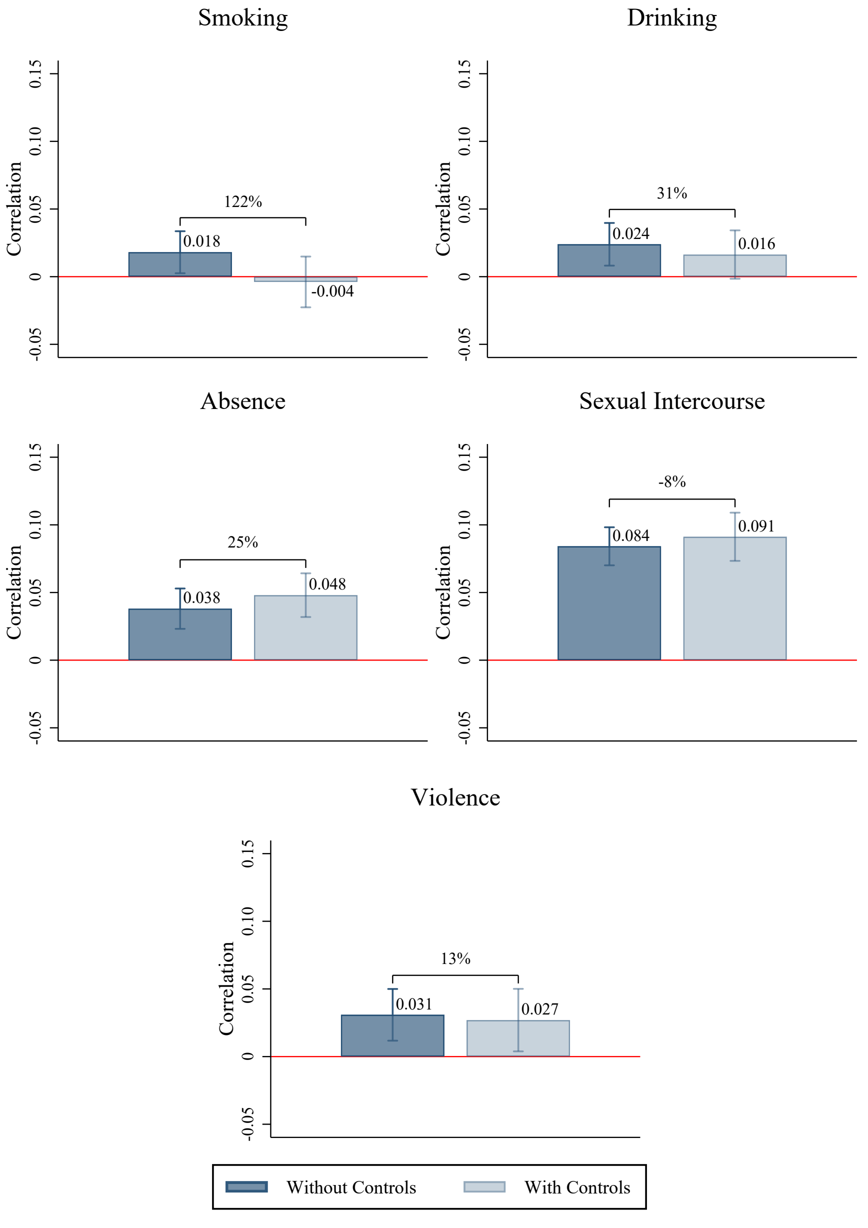

Fig. 2 Baseline Sibling Correlations in Risky Behavior (Full Sample)

Notes. Bar indicates the estimated sibling correlations (ρ) for each risky behavior. Dark bars correspond to estimates without controls, while light bars correspond to estimates with socioeconomic and demographic controls. Vertical lines represent 95% confidence intervals based on robust standard errors. The percentages above each pair of bars report the proportional change in the correlation after including controls, calculated as

$$\%\Delta\rho = \frac{\rho_{\text{without controls}} - \rho_{\text{with controls}}}{\rho_{\text{without controls}}} \times 100$$

3.1.3 Heterogeneity by Gender Composition

Figure 3 presents the sibling correlation coefficients in risky behaviors by sibling gender composition and illustrates how they change before and after controlling for parents' SES.

For brothers, sibling similarity in smoking rises from 0.035 to 0.056, in school absence from 0.081 to 0.110, and in sexual intercourse from 0.099 to 0.159 after controlling for parents' SES, corresponding to increases of 58%, 36%, and 60%, respectively. This suggests that parents' SES may have functioned as a suppressor. In other words, parents' SES weakened sibling similarity, and once it was controlled for, sibling similarity became more pronounced. Two mechanisms may account for this result. First, parents with higher SES may impose more individualized expectations on each child and respect autonomous choices, leading to greater behavioral divergence between siblings. In this case, parents' SES reduces similarity. Second, once the effect of parents' SES is controlled for, imitation or behavioral transmission between siblings becomes more salient, reflecting the social influence of the older brother as a role model. It should be noted that this interpretation is not causal but contextual. The coefficient change should be viewed as a plausible explanation rather than direct evidence of the older sibling's influence. By contrast, for violence, sibling similarity decreases after controlling for parents' SES, suggesting that parents' SES may serve as a protective factor. Parents with higher education levels, for example, may provide consistent norms regarding violence across all children. Among younger-brother pairs with an older sister, the correlation for school absence changes from 0.025 before controlling for SES to –0.084 after, and for sexual intercourse from 0.026 to –0.030. However, the wide confidence intervals and the relatively small sample size in this subgroup suggest caution in interpreting these sign reversals, as they may partly reflect sampling variability or low base rates of these behaviors. Nevertheless, if taken at face value, this implies that sibling similarity may not be shaped by parents' SES but rather by unobserved factors such as gender interactions, social norms, or cultural expectations. In this sense, controlling for parents' SES may have revealed the influence of these latent factors more clearly.

The results in figure 4 differ when comparing cases where the younger sibling is a girl. For sisters, sibling correlations remain generally stable, with little change. The exception is drinking, where the correlation decreases slightly after controlling for parents' SES, suggesting that parents' SES exerts some protective effect by reinforcing temperance norms or monitoring behaviors. For violence, the correlation is nearly zero with large standard errors, indicating no explanatory power. These results suggest that parents' SES exerts a relatively limited influence on sisters' behaviors. In contrast, when the older sibling is a brother and the younger is a sister, sibling similarity decreases in most risky behaviors except absence. The decline is particularly pronounced for drinking, suggesting that parents' SES may have a deterrent effect by mediating behavioral transmission. In this context, where the older brother serves as a strong social model for the younger sister, parental intervention may have curtailed this transmission.

In sum, these findings underscore that parents' SES exerts heterogeneous effects on sibling similarity in risky behaviors, depending on sibling gender composition.

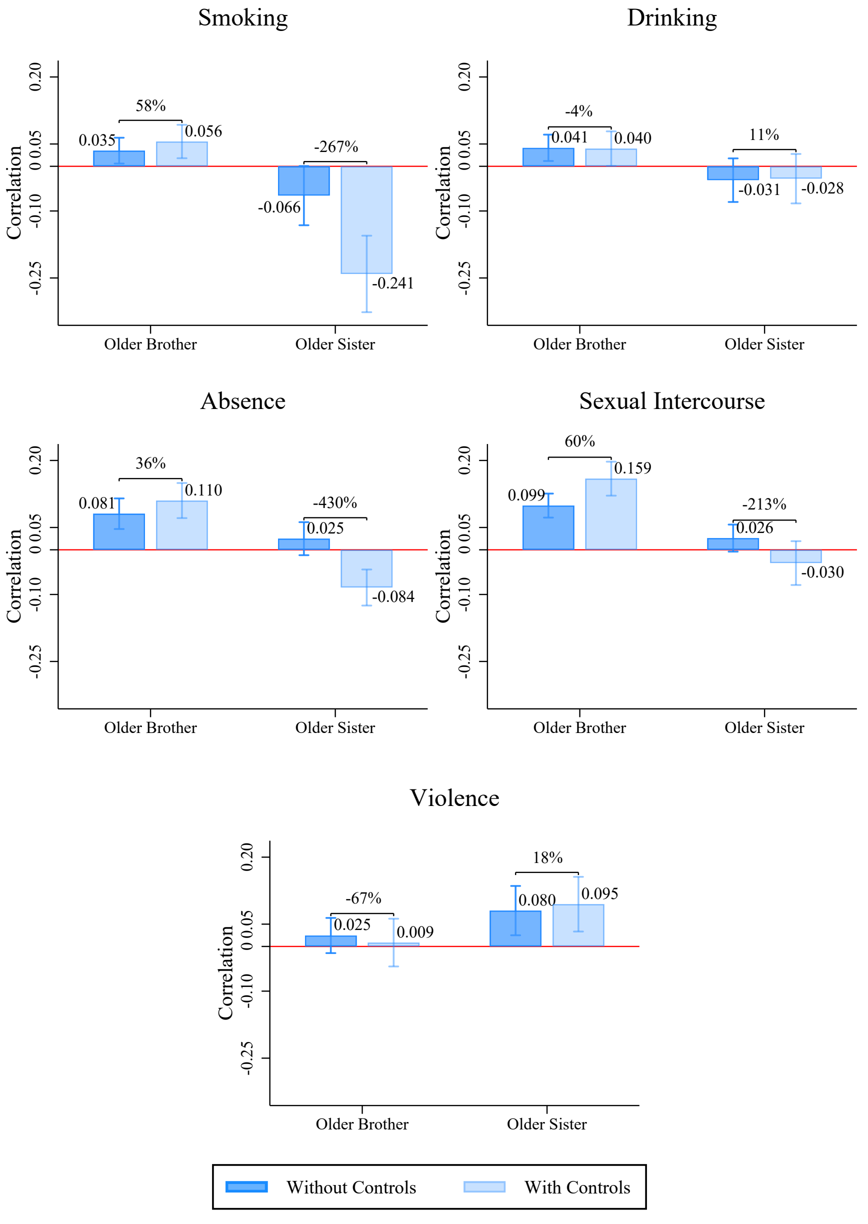

Fig. 3 Sibling Correlations in Risky Behavior by Younger Sibling's Gender: Boy

Notes. Bar indicates the estimated sibling correlations (ρ) for each risky behavior. Dark bars correspond to estimates without controls, while light bars correspond to estimates with socioeconomic and demographic controls. Vertical lines represent 95% confidence intervals based on robust standard errors. The percentages above each pair of bars report the proportional change in the correlation after including controls, calculated as

$$\%\Delta\rho = \frac{\rho_{\text{without controls}} - \rho_{\text{with controls}}}{\rho_{\text{without controls}}} \times 100$$

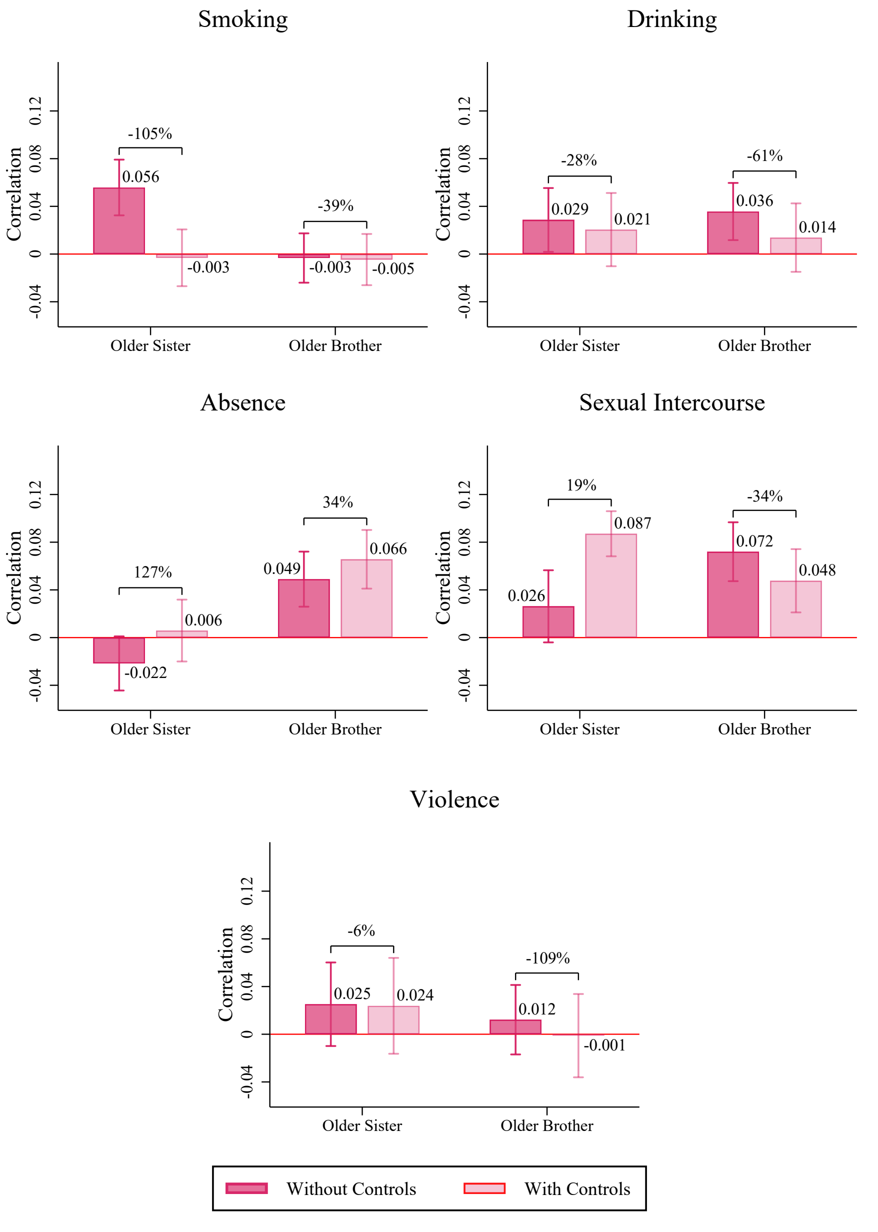

Fig. 4 Sibling Correlations in Risky Behavior by Younger Sibling's Gender: Girl
Notes. Bar indicates the estimated sibling correlations (ρ) for each risky behavior. Dark bars correspond to estimates without controls, while light bars correspond to estimates with socioeconomic and demographic controls. Vertical lines represent 95% confidence intervals based on robust standard errors. The percentages above each pair of bars report the proportional change in the correlation after including controls, calculated as

$$\%\Delta\rho = \frac{\rho_{\text{without controls}} - \rho_{\text{with controls}}}{\rho_{\text{without controls}}} \times 100$$

Gender composition should be viewed not merely as a control variable but as a factor shaping behavioral formation through sibling roles and family context Dahl et al., 2024. Controlling for parents' SES tends to accentuate sibling similarity when brothers are present, whereas the effect often weakens or becomes distorted with sisters. Overall, sibling similarity is driven less by parents' SES than by the interaction between sibling roles and the broader family environment.

3.1.4 Heterogeneity by Age Gap

Figure 5 shows how sibling correlations vary by age gap before and after controlling for parents' SES. The results indicate that despite controlling for parents' SES, correlations remain strong among siblings less than two years apart. This suggests that sibling similarity in risky behavior is more pronounced when the age difference is small, and parents' SES may function more clearly as a moderating factor. By behavior, in the group with an age difference of two years or less, smoking increases by 77%, and drinking by 50% after controlling for parents' SES. These patterns indicate that parents' SES may operate as a suppressor effect, obscuring the true extent of sibling similarity in certain behaviors. By contrast, for sexual intercourse, the effect is only -2%, indicating no evidence that SES contributes to the observed sibling correlation.

In the group with age gaps greater than two years, extreme values emerge, for instance, 120% for smoking, –149% for drinking, and –93% for sexual intercourse. This likely reflects that siblings with larger age gaps may have grown up under different parental economic conditions. Since the analysis controls for parents' SES as a fixed variable at one point, such discrepancies may arise from mismatches in the timing of household resources. In other words, the moderating role of parents' SES may be applied inconsistently, producing measurement bias.

Regardless of age difference, parents' SES partly explains violence, but the effect is limited. Sibling similarity in violent behavior likely depends more on unobserved family-level factors such as older siblings' peer influence, household discipline, or experiences of conflict. At the same time, in some cases, parents' SES may exert a deterrent effect by directly constraining the transmission of violence across siblings.

Sibling correlations remain high in the small age-gap group even after controlling for parents' SES, suggesting that the interaction between the nature of risky behaviors and age proximity may amplify the parents' SES effect. This pattern is consistent with the finding of stronger spillover effects in risky behaviors among closely spaced siblings by Argys et al. (2006). Furthermore, the results show that household resources do not moderate all risky behaviors uniformly but operate selectively by behavior. When siblings are close in age, they are more likely to attend the same school and spend more time together, making smoking, drinking, and absence particularly sensitive to both peer influence and parental control. By contrast, sexual intercourse and violence provide fewer opportunities for shared experiences, and parental intervention may be more limited, leaving unobserved traits and parenting styles to play a stronger role.

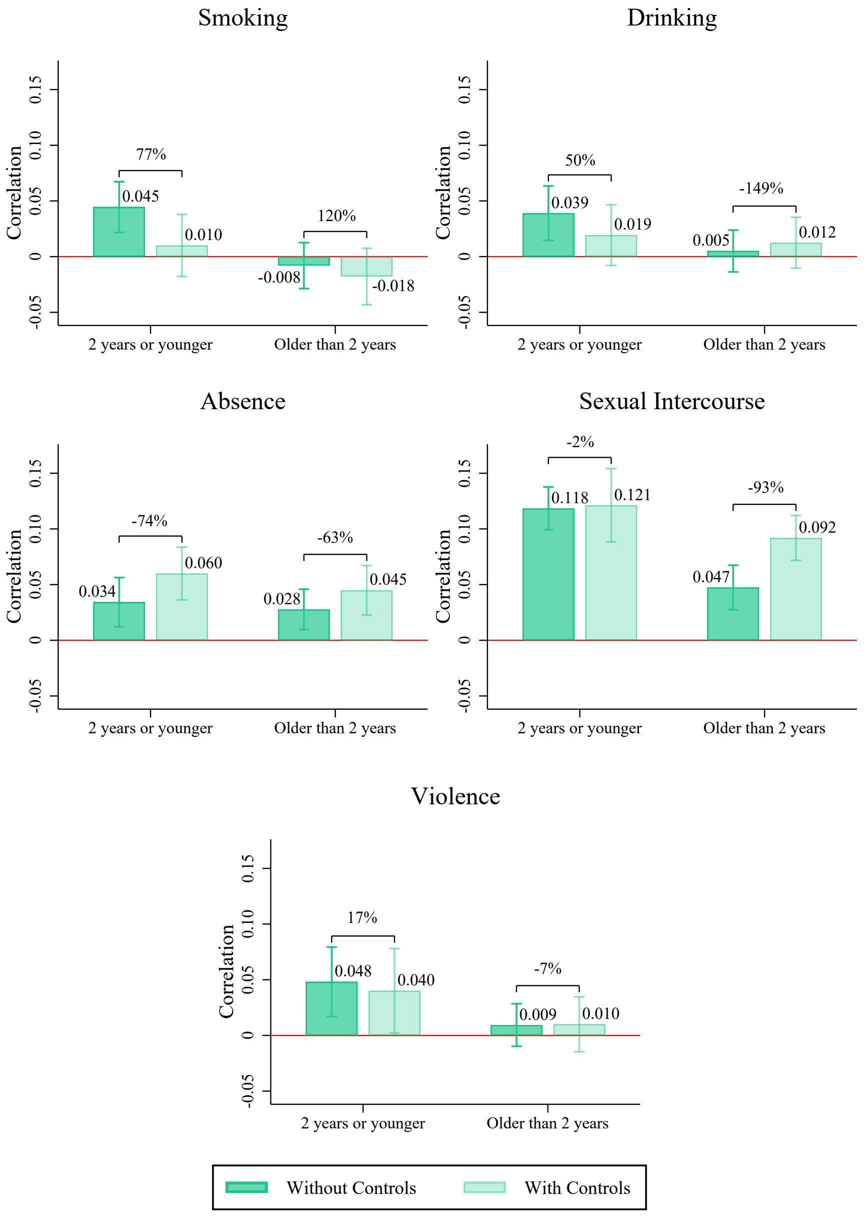

Fig. 5 Sibling Correlations in Risky Behavior by Age Gap
Notes. Bar indicates the estimated sibling correlations (ρ) for each risky behavior. Dark bars correspond to estimates without controls, while light bars correspond to estimates with socioeconomic and demographic controls. Vertical lines represent 95% confidence intervals based on robust standard errors. The percentages above each pair of bars report the proportional change in the correlation after including controls, calculated as

$$\%\Delta\rho = \frac{\rho_{\text{without controls}} - \rho_{\text{with controls}}}{\rho_{\text{without controls}}} \times 100$$

3.2 Older Sibling Effects on Younger Sibling

3.2.1 Empirical Strategy: Joint Dynamic Probit Model

In the previous decomposition analysis, I confirmed that parents' socioeconomic status can explain part of the sibling correlation in risky behaviors. However, even after controlling for family structure variables such as sibling gender composition and age gap, the correlation remains elevated, indicating that some unexplained similarity between siblings persists. This unexplained similarity may not simply arise from a shared family environment. However, it could instead reflect either a peer effect from the older sibling's behavior on the younger sibling or unobserved family-level heterogeneity. Moreover, the decomposition analysis only demonstrates correlation and cannot identify whether the residual similarity represents a causal effect. For this reason, this section investigates the structural mechanisms through which the older sibling's past and current behaviors affect the younger sibling's current behavior. To this end, I employ a Joint Dynamic Probit Model, estimated using maximum likelihood estimation (MLE).

According to Appendix Figure A1, the empirical framework is designed to capture the causal influence of the older sibling on the younger sibling, while abstracting from the reverse channel. This specification is consistent with prior research. For example, Oettinger (2000) restricts attention to the older-to-younger pathway, and Altonji et al. (2017) emphasizes that including it leads to unstable coefficients and inflated standard errors. In line with this literature, I assume that the behavior of the younger sibling does not affect the older sibling, thereby constraining the peer effect to operate in a unidirectional manner.

Siblings make binary choices regarding whether to engage in risky behavior at each period. Each individual is assumed to possess a latent utility associated with the decision to engage in such behavior. When the latent utility exceeds a certain threshold, the individual is modeled as choosing risky behaviors. Focusing first on the older sibling, the risky behavior decision at time t can be represented by the following latent utility function:

$$V_t^o = \beta_1 R_{t-1}^o + \beta_2 \mathbf{X}^o + \beta_3 \tau^o + \varepsilon_t^o \tag{5}$$

where V_t^o represents the latent utility associated with the risky behavior choice of the older sibling o at time t. R_{t-1}^o is an indicator variable that denotes whether the older sibling engaged in risky behavior at time t-1, capturing the dynamic effect of past behavior on current decision-making $\mathbf{X}^o$ is a vector of observed individual and household characteristics, including the older sibling's age, gender, parental age and education level, number of children in the household, and regional characteristics. Additionally, τ^o represents time-varying individual-specific factors, such as satisfaction with academic performance, school life, peer relationships, and parental relationships. Since these covariates do not fully capture all behavioral determinants, the residual component is subsumed into ε_t^o, assumed to be independently and identically distributed (i.i.d.) over time within an individual.

Then the older sibling engages in risky behavior if the latent utility derived from this choice exceeds the utility of not engaging in such behavior, which is normalized to zero. Accordingly, the observable risky behavior choice of the older sibling, R_t^o, is defined as follows:

$$R_t^o = 1(\beta_1 R_{t-1}^o + \beta_2 \mathbf{X}^o + \beta_3 \tau^o + \varepsilon_t^o > 0) \tag{6}$$

The younger sibling y has a latent utility function specified in an analogous form to that of the older sibling. The error term ε_t^y is defined analogously and is assumed to be i.i.d. over time within the individual. The key distinction, however, unlike the older sibling's equation, the younger sibling's latent utility explicitly includes both the current R_t^o and past R_{t-1}^o behaviors of the older sibling, as shown in equation:

$$V_t^y = \phi_1 R_t^o + \phi_2 R_{t-1}^o + \beta_1 R_{t-1}^y + \beta_2 \mathbf{X}^y + \beta_3 \tau^y + \varepsilon_t^y \tag{7}$$

This specification mitigates the confounding issue from shared family and environmental factors that would bias estimates if only the older sibling's contemporaneous behavior were included. Parental attitudes, household economic conditions, and community characteristics can affect both siblings simultaneously, potentially leading to upward bias in the estimated association between the older sibling's current behavior and the younger sibling's choice. By conditioning on both contemporaneous and lagged measures of the older sibling's behavior, the model distinguishes spurious correlations due to shared environments from the effect of sibling behavior. The observed risky behavior of the younger sibling, R_t^y, is therefore defined as follows:

$$R_t^y = 1(\phi_1 R_t^o + \phi_2 R_{t-1}^o + \beta_1 R_{t-1}^y + \beta_2 \mathbf{X}^y + \beta_3 \tau^y + \varepsilon_t^y > 0) \tag{8}$$

In equation (8), the key parameters of interest are ϕ_1 and ϕ_2, which capture two distinct channels through which the older sibling's behavior influences the younger sibling's behavioral choice. The parameter ϕ_1 represents the short-run, direct effect of the older sibling's contemporaneous behavior on the younger sibling's decision. This effect can be interpreted as an imitation mechanism, whereby repeated exposure to the older sibling's current actions increases the probability that the younger sibling also engages in risky behavior. A positive estimate of ϕ_1 ($\phi_1 > 0$) thus reflects a dynamic interaction effect, directly related to the reflection problem highlighted by Manski (1993). To address this simultaneity, the model imposes a causal ordering based on age, assuming that the older sibling may influence the younger sibling but not the reverse. By contrast, ϕ_2 captures the indirect, longer-run influence of the older sibling's past behavior on the younger sibling's current decision. This channel reflects behavioral persistence and the diffusion of cultural or normative factors within the household. Specifically, the younger sibling may form expectations or internalize household norms based on the older sibling's prior behavior. Hence, a positive estimate of ϕ_2 ($\phi_2 > 0$) indicates not merely a contemporaneous correlation, but a long-term transmission of behavior. This distinction aligns with Altonji et al. (2017) emphasis on separating peer effects from confounding family background factors.

When estimating heterogeneity by sibling gender composition or age gap, the dynamic joint probit model encounters convergence difficulties arising from sample size reductions. To mitigate these issues, I additionally implement a fixed effects model based on the Frisch–Waugh–Lovell theorem. Although the FE model employs a linear probability specification rather than the nonlinear probit, it serves as a robust alternative that identifies the same within-household variation while avoiding convergence problems in smaller subsamples.

Finally, the peer effect of the older sibling in this study is restricted to within-household channels, and indirect influences through external peer networks or community-level interactions are excluded. There are two main reasons for this restriction. First, there is an identification concern. External social networks, such as friendships, are likely to correlate highly with the older sibling's behavior. Suppose the younger sibling interacts with the older sibling's peers. In that case, it becomes difficult to distinguish whether changes in the younger sibling's behavior are due to the older sibling's direct influence or the indirect spillover through peers. This omitted variable problem generates endogeneity, as unobserved external factors correlate with R_{ot-1}, thereby increasing the estimated peer effect. Second, there are issues of simultaneity and selection bias (Manski, 1993). In external networks, the younger sibling and his or her peers may influence each other's behavior simultaneously, preventing a clear identification of causal direction. Moreover, peer formation is often subject to assortative matching, whereby the younger sibling may choose friends with similar inherent traits or predispositions. In this case, observed behavioral similarity would reflect initial conditions or unobserved preferences, rather than a causal peer effect. Given these challenges, incorporating external peer networks would not resolve the problem but rather exacerbate endogeneity concerns. Therefore, the analysis focuses exclusively on direct within-household peer effects of the older sibling, leaving the role of external social networks as an important question for future research.

3.2.2 Older Sibling Peer Effect Over Time, Past and Present

Table 2 shows estimates from the full sample on the effect of older siblings' contemporaneous and lagged behaviors on younger siblings' risky behaviors. The results reveal three salient patterns.

First, contemporaneous effects dominate. Across all risky behaviors, the older sibling's current behavior exerts a more decisive influence than past behavior. In several cases, contemporaneous effects are statistically significant, whereas lagged effects are generally insignificant. This suggests that sibling spillovers decay rapidly over time. Second, there is substantial heterogeneity across types of risky behavior. For smoking and sexual intercourse, neither contemporaneous nor lagged effects are statistically significant, indicating no evidence of sibling influence. By contrast, drinking shows a contemporaneous estimate of about 0.64, statistically significant at the 5 percent level. Absence shows an effect of approximately 1.03, which is also significant at the 5 percent level. Violence displays an estimate of about 0.40, which is significant at the 1 percent level. These findings indicate that sibling transmission operates selectively, with more potent effects in domains such as drinking, absence, and violence, where social exposure and opportunities for sharing are more prominent.

Table 2 Contemporaneous and Lagged Effects of Older Siblings

	Smoking	Drinking	Absence	Sexual Intercourse	Violence
R_t^o	0.551	0.639**	1.026**	1.079	0.400***
	(0.795)	(0.250)	(0.401)	(0.745)	(0.140)
R_{t-1}^o	0.320	-0.176	0.280	0.974	-0.327
	(0.556)	(0.327)	(0.580)	(0.829)	(0.377)
R_{t-1}^y	2.997***	2.153***	2.483***	1.404**	2.214***
	(0.264)	(0.181)	(0.229)	(0.562)	(0.132)
N	9,156	9,156	9,156	9,156	9,156

Notes. Robust standard errors clustered at the individual level, with significance denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

Third, the role of the older sibling is relatively modest compared to state dependence. Although the primary focus of this study is on older sibling spillover, the estimates show that younger siblings' risky behaviors are strongly determined by their past choices. This underscores that the persistence of risky behavior within individuals explains more variation than sibling influence. Similarly, Altonji et al. (2017) report that older siblings' behaviors provide important signals for younger siblings' initial choices, with more potent effects observed at earlier stages than in later periods. While the present study does not explicitly estimate the timing of initial onset, the strong predictive power of younger siblings' lagged behaviors suggests that early intervention is critical from a policy perspective. If prevention fails at the initial stage, subsequent persistence is likely to dominate.

Finally, it is important to emphasize that the estimated sibling effect coefficients may capture not only causal influences but also transitory shocks or family-level events that simultaneously affect both siblings. For this reason, the estimates should be interpreted not as precise causal effects but rather as an upper bound of the causal effect.

3.2.3 Heterogeneity by Gender Composition

Table 3 summarizes the contemporaneous and lagged effects of older siblings by gender composition. To unpack these patterns, I first turn to Panel A, which covers brother–brother pairs. Within this group, the contemporaneous effects of the older sibling are generally small and mostly insignificant. An exception arises in the case of sexual intercourse, where the coefficient is 0.168, suggesting a relatively large positive association. This finding may indicate the possibility of immediate imitation between brothers, even though this does not have a significant effect. Turning to the older sister and younger brother pairs, a consistent pattern emerges across smoking, drinking, and sexual intercourse. In these cases, the older sister's influence appears to function less through straightforward imitation and more as a suppressing signal. For smoking, the contemporaneous effect is -0.205 and statistically significant at the 5 percent level, implying that when the older sister currently smokes, the younger brother is, in fact, less likely to smoke. In drinking, the lagged effect is positive, which

Table 3 Contemporaneous and Lagged Effects of Older Siblings by Gender Composition

	Smoking	Drinking	Absence	Sexual Intercourse	Violence
Panel A. Boy					
Same Gender (Older : Boy), N = 856					
R^o_t	0.010	0.030	0.077	0.168	-0.016
	(0.033)	(0.039)	(0.093)	(0.105)	(0.040)
R^o_{t-1}	-0.037	0.016	0.006	0.06	0.035
	(0.034)	(0.033)	(0.026)	(0.099)	(0.032)
R^y_{t-1}	-0.061	-0.376***	-0.075	0.431**	-0.177*
	(0.236)	(0.137)	(0.249)	(0.167)	(0.104)
Diff Gender (Older : Girl), N = 763					
R^o_t	-0.205 **	-0.028	-0.086	-0.059	0.043
	(0.103)	(0.039)	(0.080)	(0.055)	(0.037)
R^o_{t-1}	0.093	0.112**	-0.006	-0.084 *	-0.015
	(0.078)	(0.056)	(0.017)	(0.051)	(0.029)
R^y_{t-1}	-0.156	-0.207	-0.483***	-0.589***	-0.235*
	(0.192)	(0.144)	(0.021)	(0.106)	(0.122)
Panel B. Girl					
Same Gender (Older : Girl), N = 870					
R^o_t	-0.001	-0.006	-0.013	0.102	-0.004
	(0.003)	(0.021)	(0.039)	(0.095)	(0.028)
R^o_{t-1}	0.004	-0.011	0.04	0.016	-0.034
	(0.003)	(0.041)	(0.082)	(0.018)	(0.036)
R^y_{t-1}	-0.230 ***	-0.052	-0.306	-0.242***	-0.141
	(0.013)	(0.369)	(0.219)	(0.012)	(0.115)
Diff Gender (Older : Boy), N = 873					
R^o_t	0	0.006	0.053	0.039	-0.015
	(0.001)	(0.023)	(0.049)	(0.031)	(0.019)
R^o_{t-1}	0.002	-0.037	0.058	0.077	0.013
	(0.003)	(0.031)	(0.054)	(0.075)	(0.020)
R^y_{t-1}	0.316	-0.151	-0.252***	-0.309***	-0.207 **
	(0.253)	(0.209)	(0.005)	(0.053)	(0.089)

Notes. Robust standard errors clustered at the individual level, with significance denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

shows 0.112, significant at the 5 percent level, suggesting that the younger brother may initially imitate the older sister's past drinking behavior. However, the contemporaneous coefficient is -0.028, pointing to a possible transition from imitation to suppression over time. For sexual intercourse, the contemporaneous coefficient is -0.059, and the lagged coefficient, -0.084, is negative and marginally significant at the 10 percent level. These results suggest that the older sister's sexual intercourse experience lowers the probability that the younger brother engages in sexual intercourse. Taken together, these results suggest an immediate suppressing effect in smoking, a shift from initial imitation to suppression in drinking, and a consistently suppressing effect in sexual intercourse. In Panel B, I find no significant effect of older sibling effects in either sister pairs or older brother and younger sister pairs.

In general, the sibling effects by gender composition reveal no clear consistency. Prior studies often find stronger behavioral transmission within same gender dyads due to homophily, but in this study, no significant contemporaneous effects appear in same gender pairs. Instead, in some domains of risky behavior, I observe negative effects within mixed gender pairs, particularly between older sisters and younger brothers. This implies that the older sister's behavior may operate less as a template to emulate and more as a point of differentiation or regulation for the younger brother. This finding highlights that sibling effects by gender composition may not be confined to simple imitation patterns. However, the standard errors of the gender specific estimates are relatively large, which makes it difficult to regard the observed negative effects or suppressing patterns as statistically robust. Therefore, the findings of this study should be interpreted as suggestive evidence that imitation and suppression effects may manifest heterogeneously depending on sibling gender composition.

3.2.4 Heterogeneity by Age Gap

Table 4 shows that the effects of older siblings on younger siblings' engagement in risky behavior, examined by age gap, are not statistically significant. Thus, the null hypothesis of no effect cannot be rejected. Nonetheless, the direction and size of the estimates suggest an interesting pattern. Younger siblings appear more strongly influenced when the sibling age gap is two years or less, as shown in Panel A, compared to when it is three years or more, as shown in Panel B.

When the age gap is narrow, younger siblings tend to imitate, at least to some extent, the contemporaneous behaviors of their older siblings, while residual influences of past behaviors also remain. For instance, the younger sibling's prior behavior significantly decreases the probability of repeating drinking and school absence by 30 percent and 41 percent, respectively. This suggests that prior engagement in such behaviors operates as a suppressing rather than a reinforcing mechanism, pointing to weak self-persistence among younger siblings. Consequently, the current or past behaviors of older siblings may serve as relatively more salient explanatory factors in the behavioral choices of younger siblings. Consistent with this interpretation, many coefficients in Panel A are close to zero or positive, suggesting a weak but present spillover effect of contemporaneous older sibling behaviors.

By contrast, in Panel B, the coefficients are more irregular in both sign and size

Table 4 Contemporaneous and Lagged Effects of Older Siblings by Age Gap

	Smoking	Drinking	Absence	Sexual Intercourse	Violence
Panel A. Two years or less apart, N = 1,733					
R_t^o	-0.001	0.017	0.036	0.135	0.005
	(0.025)	(0.027)	(0.049)	(0.106)	(0.026)
R_{t-1}^o	-0.003	0.029	0.071	0.03	0.008
	(0.025)	(0.026)	(0.052)	(0.138)	(0.023)
R_{t-1}^y	-0.039	-0.306**	-0.413***	0.177	-0.159 **
	(0.196)	(0.124)	(0.083)	(0.241)	(0.068)
Panel B. More than two years apart, N =1,629					
R_t^o	-0.023	-0.002	0.009	0.074	-0.010
	(0.017)	(0.016)	(0.048)	(0.060)	(0.020)
R_{t-1}^o	0.014	-0.024	0.009	0.053	0.002
	(0.014)	(0.029)	(0.025)	(0.042)	(0.019)
R_{t-1}^y	-0.012	-0.134	-0.002	-0.280 *	-0.363 ***
	(0.184)	(0.159)	(0.244)	(0.153)	(0.085)

Notes. Robust standard errors clustered at the individual level, with significance denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

and lack statistical significance. This implies that as the age gap widens, the influence of older siblings' behaviors diminishes or, at the very least, plays a less substantial role in shaping the younger sibling's behavioral decisions.

When comparing specific behaviors, I find that, except for smoking, most risky behaviors display larger and more positive contemporaneous and lagged effects in Panel A than in Panel B. For example, in drinking, both past and present coefficients are positive in Panel A but turn negative in Panel B. Similarly, sexual intercourse shows a positive contemporaneous coefficient of 0.135 in Panel A, whereas in Panel B it shifts to a negative coefficient of -0.280. These patterns suggest that close age spacing fosters imitation and spillover, whereas wider spacing weakens or even reverses such dynamics.

In summary, when siblings are close in age, younger siblings are more likely to imitate older siblings' behaviors in the short term and experience spillover effects in the longer term. By contrast, as the age gap increases, these effects weaken or shift in a negative direction. This finding is consistent with the broader result that the contemporaneous influence of older siblings is generally stronger than their lagged influence, supporting the interpretation that spillover effects fade over time. However, given that the reduced sample size increases standard errors and lowers the precision of the estimates, the statistical significance of these findings is limited. Interpretation should therefore focus on the direction and relative magnitudes of the estimates rather than their exact significance levels.

3.3 Unobserved Common Factors Influencing Siblings

3.3.1 Empirical Strategy: Correlated Random Model

Up to this point, the analysis has examined the explanatory power of parents' SES for sibling behavioral similarity, as well as the direction and magnitude of the older sibling's effect. However, these factors alone cannot fully account for the observed similarity, suggesting that unobserved common factors may also be at work. To address this residual correlation, I use the correlated random effects (CRE) model introduced by Altonji et al. (2017). The purpose of this approach is to distinguish the peer effect of the older sibling on the younger sibling from persistent shared environmental influences, such as parental attitudes and disciplinary practices. The model can be written as:

$$R_t^y = \lambda_0 + \lambda_1 (R_{t-1}^o + R_t^o + R_{t+1}^o) + \lambda_2 R_t^o + \lambda_3 R_{t-1}^o + \mathbf{X}^y + \lambda_4 \tau^y + \varepsilon \qquad (9)$$

The variables used are identical to those in equation (8), with the addition of $(R_{t-1}^o + R_t^o + R_{t+1}^o)$, which is called a common factor. This variable captures the possibility that unobserved family environments, such as parental attitudes and disciplinary practices, operate not as temporary shocks but as persistent and consistent influences on both children over time. It functions as a Mundlack-type correction, accounting for unobserved family-level heterogeneity rather than representing a causal parameter. Including a common factor variable for the older sibling is motivated by the idea that if similar behavioral patterns emerge across adjacent periods, this provides strong evidence that the parental environment exerts a continuous and consistent influence. So, a common factor serves to identify long-term shared environmental effects. If a common factor is correlated with ε_t^y, this correlation is interpreted not as mere imitation but as the result of unobserved shared family factors such as parental discipline and child-rearing practices. Accordingly, the parameter of interest is λ_1, which captures unobserved shared family influences. Namely, the common factor, while λ_2 and λ_3 represent the older sibling's contemporaneous and lagged peer effects, respectively. Under this specification, λ_1 accounts for shared environmental effects correlated with the error term, allowing λ_2 and λ_3 to be interpreted as pure peer effects.

Following Altonji et al. (2017), identification of the correlated random-effects model relies on three key assumptions. First, the unobserved common factor that represents shared family influences, such as parenting norms, disciplinary practices, or household values, is assumed to affect both siblings comparably. Parents are expected to apply similar rules and expectations to their children. For instance, they might impose the same curfews or academic standards on both. This symmetry ensures that parental attitudes operate similarly across siblings. Second, the younger sibling's past behavior should not directly determine the current outcome, and the model assumes no serial correlation in the error term, such that $E[\varepsilon_t^y \mid R_{t-1}^y] = 0$ *and* $\text{Cov}(\varepsilon_t^y, \varepsilon_{t-1}^y) = 0$. If the younger sibling's behavior is highly persistent, it becomes difficult to separate the influence of the older sibling from the persistence of the younger sibling's own trajectory. For example, if a younger sibling repeatedly skips school across years, this pattern may reflect self-continuing behavior rather than genuine sibling influence; such cases are therefore excluded. Third, idiosyncratic shocks to the older sibling are assumed not to persist over time. Otherwise, the coefficient λ_1 would confound family-level effects with sibling-specific disturbances, complicating interpretation.

3.3.2 The Influence of Unobserved Common Factors

Table 5 reports the results of equation (9) for each type of risky behavior using the full sample.

Beginning with smoking, all factors are statistically significant, and the estimated magnitudes are small. This is consistent with the earlier finding from the dynamic joint probit model that sibling effects on smoking are weak. The results suggest that variation in smoking is driven mainly by idiosyncratic shocks at the individual level, with little correlation between common factors and the error term. In other words, the variance explained by λ_1 is limited, implying that the identified effects are weak and that smoking may be more strongly influenced by individual choices related to utility maximization rather than family environment.[12]

By contrast, drinking is clearly shaped by common factors. The effect of the common factor is 8 percent, and it is statistically significant at the 1% level, suggesting that parental attitudes and monitoring practices jointly influence the drinking decisions of both siblings. Interestingly, the contemporaneous effect of the older brother's drinking behavior is negative and statistically significant at the 5% level, which can be interpreted as a suppressing rather than imitative effect. In other words, the older sibling's drinking may be perceived by the younger sibling as a risk factor, thereby functioning as a protective mechanism that discourages drinking. Considering both the magnitude and direction of the effects, drinking behavior appears to be shaped primarily by family-level social and psychological interactions and persistent household factors (Balsa et al., 2018).

For school absence, all factors exhibit positive signs but are not statistically significant. It suggests that the correlation between explanatory variables and the error term is weak, and that the within-family variation itself is limited, leading to significant standard errors and making identification of effects difficult. In particular, strong external institutional constraints, such as South Korea's compulsory education system, may dominate school attendance decisions, thereby reducing the relative contribution of family and sibling influences.

Sexual intercourse exhibits a pattern distinct from the other types of risky behaviors. The lagged effect of the older sibling is significant, with an estimated magnitude of about 8 percent, whereas the contemporaneous effect is not statistically significant. This suggests that the older sibling's sexual intercourse behavior is not immediately correlated with the younger sibling's error term at the time of occurrence but instead is transmitted over time as a family norm that eventually influences the younger sibling's behavior when the common factor is controlled. Processes of information

[12] According to the Korea Youth Risk Behavior Survey conducted by the Korea Disease Control and Prevention Agency, 52.1% of middle school student, 50.4% of high school student reported that curiosity was the primary reason for initiating smoking. (Korea Disease Control and Prevention Agency (2021))

Table 5 The Influence of Unobserved Common Factors

	Smoking	Drinking	Absence	Sexual Intercourse	Violence
$R^o_{t-1}+R^o_t+R^o_{t+1}$	0.016	0.083***	0.039	0.027	0.101***
	(0.018)	(0.022)	(0.037)	(0.040)	(0.027)
R^o_t	-0.016	-0.050**	0.034	0.098	-0.030
	(0.035)	(0.025)	(0.047)	(0.076)	(0.030)
R^o_{t-1}	0.002	-0.015	0.033	0.089**	-0.056*
	(0.025)	(0.034)	(0.049)	(0.044)	(0.028)
N	2,720	2,720	2,720	2,720	2,720

Notes. Robust standard errors clustered at the individual level, with significance denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

asymmetry and delayed norm transmission can explain such dynamics. It means that even if the older sibling engages in sexual intercourse contemporaneously, it is unlikely that this information might become immediately known to the younger sibling, thereby limiting direct imitation. Over time, however, this information may be shared and internalized as a family norm, ultimately shaping the younger sibling's sexual intercourse decisions in the long run. Hence, sexual intercourse appears to be influenced less by short-term imitation and more by a gradual process of learning and intergenerational transmission of norms.

Lastly, violence reveals an intriguing mechanism in which shared common factor and lagged effects operate in opposite directions. The common factor effect is positive and statistically significant, indicating that persistent family-level influences play an important role in shaping the learning of violent behavior. At the same time, the coefficient on the older sibling's past violent behavior is negative and statistically significant, suggesting that the older sibling's adverse experiences generate a deterrent effect, whereby the younger sibling avoids violence through a learning pathway.

Overall, three main findings emerge. First, the determinants of younger siblings' risky behaviors differ by behavioral domain. A single mechanism cannot explain these behaviors, nor can it be assumed that siblings raised in the same household consistently share behavioral patterns. Second, the influence of older siblings generally operates as a deterrent effect when the common factor is controlled. Except for behaviors such as school absence and sexual intercourse, where institutional or private motives are more salient, the older sibling's negative experiences tend to reduce the likelihood of the same behaviors in the younger sibling through deterrence and risk-avoidance pathways. Third, directly observable and learnable behaviors within the household exhibit stronger explanatory power of a shared common factor. For instance, drinking and violent behavior are strongly shaped by parental lifestyles and disciplinary practices, which are readily exposed to children. By contrast, smoking, sexual intercourse, and school absence are less visible to parents in daily life, limiting the influence of shared environment and making these behaviors more likely to be driven by peer groups or individual-specific factors.

However, it is important to note that the estimation of common factors, λ_1, may be unstable when the overall variation in behaviors is small or within-family variation is limited. Given the panel structure, certain age groups or behaviors classified as rare events may involve relatively small sample sizes, which can inflate the standard errors of the estimated coefficients and render the results more susceptible to noise.

3.3.3 The Influence of Unobserved Common Factors by Gender Composition

In Table 6, I present the next set of results, which examines whether the factors and mechanisms shaping younger siblings' risky behaviors differ according to the sibling pair's gender composition.

For younger brothers in Panel A, the presence of an older brother implies that the common factor is the primary determinant of behavior. Specifically, drinking and violence both increase by approximately 8 percent and 13 percent, respectively, and are statistically significant at the 5% level. This indicates that common factor such as parental supervision and household atmosphere consistently influence the behaviors of both siblings. By contrast, younger brothers with an older sister exhibit a more complex pathway. A common factor significantly increases smoking and absence by 19 percent and 25 percent, respectively. At the same time, the older sister's contemporaneous smoking exerts a suppressing effect of about 47 percent. In addition, the past sexual intercourse of the older sister increases the younger brother's sexual intercourse by 8.9 percent, suggesting that in opposite-gender sibling pairs, both shared environmental influences and the older sister's contemporaneous and lagged behaviors jointly play important roles, reflecting heterogeneous mechanisms of influence.

Panel B, focusing on younger sisters, reveals patterns similar to those observed for younger brothers. Most peer effects are limited when the older sibling is a sister, with the only statistically significant increase of 12 percent in drinking attributable to a common factor. In contrast, when the older sibling is a brother, the common factor and the brother's behaviors operate simultaneously. A common factor significantly increases drinking, sexual intercourse, and violence by 7 percent, 5.9 percent, and 7.5 percent, respectively, all at the 5% significance level. At the 10% significance level, the older brother's contemporaneous drinking and sexual intercourse reduce the younger sister's behaviors by approximately 5 percent and 6 percent, respectively. In comparison, the older brother's past violent behavior reduces the younger sister's violent behavior by about 5 percent. These results suggest that the older brother's experiences with risky behaviors generate a suppressing effect on the younger sister. At the same time, the unobserved family environment reinforces the propensity toward such behaviors, indicating the coexistence of heterogeneous mechanisms of influence.

Overall, the results indicate that same-gender sibling pairs are primarily characterized by shared environmental influences, reflecting the consistent application of parental discipline and expectations. By contrast, opposite-gender sibling pairs are shaped not only by shared environment but also by more complex sibling interaction effects. Depending on the type of risky behavior, both short-run deterrent effects ($\lambda_2 < 0$) and delayed intergenerational transmission of norms ($\lambda_3 > 0$) can emerge simultaneously. These patterns are most evident in brother–sister and sister–brother pairs, where deterrent effects and shared environmental influences coexist in various forms. This suggests that parents may apply more uniform social expectations

Table 6 The Influence of Unobserved Common Factors by Gender Composition

	Smoking	Drinking	Absence	Sexual Intercourse	Violence
Panel A. Boy					
Same Gender (Older : Boy), N = 692					
$R^o_{t-1}+R^o_t+R^o_{t+1}$	-5.28E-05	0.086**	0.045	0.108	0.137**
	(0.031)	(0.039)	(0.066)	(0.097)	(0.064)
R^o_t	0.070	-0.050	0.082	0.121	-4.18E-02
	(0.050)	(0.050)	(0.078)	(0.146)	(0.074)
R^o_{t-1}	-0.023	0.030	-0.031	0.147	-0.080
	(0.063)	(0.065)	(0.079)	(0.142)	(0.075)
Diff Gender (Older : Girl), N = 616					
$R^o_{t-1}+R^o_t+R^o_{t+1}$	0.192*	0.076	0.253*	0.031	0.108
	(0.100)	(0.051)	(0.134)	(0.057)	(0.067)
R^o_t	-0.479**	-0.077	-0.179	0.036	0.023
	(0.193)	(0.096)	(0.167)	(0.049)	(0.069)
R^o_{t-1}	0.099	0.053	-0.106	0.089*	-0.051
	(0.0919	(0.070)	(0.116)	(0.046)	(0.055)
Panel B. Girl					
Same Gender (Older : Girl), N = 704					
$R^o_{t-1}+R^o_t+R^o_{t+1}$	-0.004	0.127**	-0.048	-0.043	0.048
	(0.006)	(0.059)	(0.045)	(0.040)	(0.045)
R^o_t	-0.004	-0.062	0.065	0.188	0.005
	(0.006)	(0.046)	(0.059)	(0.166)	(0.036)
R^o_{t-1}	-0.008	-0.050	0.145	-0.008	-0.036
	(0.009)	(0.072)	(0.130)	(0.018)	(0.064)
Diff Gender (Older : Boy), N = 708					
$R^o_{t-1}+R^o_t+R^o_{t+1}$	-0.007	0.071**	0.009	0.059**	0.075**
	(0.007)	(0.034)	(0.013)	(0.026)	(0.034)
R^o_t	0.002	-0.058*	0.082	-0.060*	-0.064
	(0.006)	(0.034)	(0.083)	(0.033)	(0.040)
R^o_{t-1}	0.001	-0.066	0.006	0.039	-0.056*
	(0.006)	(0.049)	(0.028)	(0.047)	(0.029)

Notes. Robust standard errors clustered at the individual level, with significance denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

and discipline to same-gender children, while expecting differentiated roles from opposite-gender children, thereby allowing the behaviors of older siblings to function more strongly as role models. Ultimately, the determinants of risky behaviors cannot be explained solely by sibling effects. Instead, they reflect the interplay of norm transmission, parental influence, and sibling interactions, which operate heterogeneously across gender composition and behavioral domains.

3.3.4 The Influence of Unobserved Common Factors by Age Gap

I investigate how the common family factor affects the mechanisms shaping younger siblings' risky behaviors, and whether its influence varies with the size of the age gap.

Table7 shows the results that in both groups, which are Panel A and Panel B, the unobserved common factor, λ_1, significantly contributes to drinking and violence. Specifically, the marginal effects of unobserved common factors on drinking were estimated at 6.8 percent in Panel A and 10.1 percent in Panel B, while the corresponding marginal effects on violence are estimated at 11.5 percent and 8 percent, respectively. These findings suggest that family environment is a key driver of sibling similarity in these two risky behaviors across age-gap groups.

When the common factor is controlled, the peer effect of the older sibling is found to be significant only through past behaviors, λ_3. In Panel A, the older sibling's past violence reduces the younger sibling's violence by 8 percent at the 5% significance level. In contrast, in Panel B, the older sibling's past sexual intercourse increases the younger sibling's sexual intercourse by 6 percent at the 10% significance level. In both cases, the contemporaneous behaviors of the older sibling λ_2 are not statistically significant. In contrast, the lagged behaviors, λ_3, are suggesting that the transmission operates less through short-term imitation than through cumulative learning and the intergenerational transmission of norms over time. Specifically, among closely spaced siblings, the older sibling's past violence appears to generate a deterrent effect, reducing the propensity for violent behavior in the younger sibling. By contrast, when the age gap is larger, information asymmetry may give rise to a delayed effect, whereby only sexual intercourse is transmitted over time through the intergenerational transmission of norms.

The tendency for significant effects to emerge only from the older sibling's past behaviors, rather than contemporaneous behaviors, suggests that sibling similarity is not merely the result of simultaneous correlation or common shocks after accounting for the common factor. Instead, it points to the operation of a learning and norm-formation channel through which the older sibling's experiences are transmitted to the younger sibling over time. Importantly, this pattern can be interpreted as evidence that alleviates concerns regarding the reflection problem, particularly the difficulty of identifying endogenous effects as highlighted by Manski (1993). If sibling behaviors were determined simultaneously in a mutually dependent structure, the contemporaneous term, λ_2, should have been significant. The fact that only the lagged term, λ_3, is significant instead provides stronger support for delayed intergenerational transmission of norms rather than contemporaneous endogeneity. In other words, these results offer circumstantial evidence consistent with the notion that changes in older siblings' behaviors causally influence the behavioral formation of younger siblings, thereby lending empirical support to the identifying assumption that peer effects operate from older to younger siblings.

Table 7 The Influence of Unobserved Common Factors by Age Gap

	Smoking	Drinking	Absence	Sexual Intercourse	Violence
Panel A. Two years or less apart, N = 1,408					
$R^o_{t-1}+R^o_t+R^o_{t+1}$	0.003	0.068***	-0.004	-0.042	0.115***
	(0.026)	(0.026)	(0.074)	(0.108)	(0.039)
R^o_t	0.003	-0.049	0.058	0.224	-0.050
	(0.056)	(0.035)	(0.103)	(0.250)	(0.049)
R^o_{t-1}	0.009	0.001	0.120	0.228	-0.090**
	(0.045)	(0.036)	(0.143)	(0.143)	(0.043)
Panel B. More than two years apart, N =1,312					
$R^o_{t-1}+R^o_t+R^o_{t+1}$	0.021	0.101***	0.059	-0.009	0.080**
	(0.023)	(0.035)	(0.043)	(0.028)	(0.037)
R^o_t	-0.023	-0.055	0.021	0.089	-0.005
	(0.041)	(0.036)	(0.043)	(0.056)	(0.033)
R^o_{t-1}	0.006	-0.026	-0.004	0.066*	-0.023
	(0.019)	(0.057)	(0.020)	(0.037)	(0.037)

Notes. Robust standard errors clustered at the individual level, with significance denoted by *, **, and *** for the 10, 5, and 1 percent levels, respectively.

3.4 Structural Decomposition on Sibling Similarity

3.4.1 Empirical Strategy: Structural Decomposition

Up to this point, I have examined how parents' socioeconomic status explains sibling correlations, the causal influence of older siblings' risky behaviors, and the role of unobserved family factors shared within the household. While these analyses yield important insights into each mechanism, it does not, on their own, reveal the relative contributions of these three components to the overall similarity in siblings' behaviors. As the final stage of the analysis, I develop a structural decomposition framework that quantitatively partitions the observed sibling correlation into these components to understand better the relative importance of socioeconomic resources, direct peer influences, and shared family environments in shaping behavioral similarity between siblings.

To maintain consistency, the younger sibling's risky behavior is specified as a function of parents' socioeconomic status, the older sibling's behavior, and common unobserved family characteristics, together with an idiosyncratic error term for the younger sibling:

$$R^y = \gamma R^o + \mathbf{X}\theta + \alpha + u^y \tag{10}$$

where R^y and R^o denote the younger and older siblings' risky behavior, $\mathbf{X}$ is a vector of parents' socioeconomic status covariates, α represents unobserved family common factors shared within the household, and u^y is an idiosyncratic shock. From this specification, the covariance between the siblings' outcomes can be written as

$$\mathrm{Cov}(R^y, R^o) = \gamma \mathrm{Var}(R^o) + \mathrm{Cov}(\mathbf{X}\theta, R^o) + \mathrm{Cov}(\alpha, R^o) \tag{11}$$

Under the assumption that the idiosyncratic error is orthogonal to the older sibling's behavior, $\mathrm{Cov}(u^y, R^o) = 0$. Accordingly, the sibling correlation is defined as

$$\rho = \frac{\gamma \mathrm{Var}(R^o) + \mathrm{Cov}(\mathbf{X}\theta, R^o) + \mathrm{Cov}(\alpha, R^o)}{\sqrt{\mathrm{Var}(R^y)\,\mathrm{Var}(R^o)}} \tag{12}$$

This decomposition highlights three interpretable components of sibling similarity. The term $\gamma \mathrm{Var}(R^o)$ captures the causal effect of the older sibling's behavior, $\mathrm{Cov}(\mathbf{X}\theta, R^o)$ reflects the contribution of parents' socioeconomic status, and $\mathrm{Cov}(\alpha, R^o)$ represents the influence of unobserved common factors jointly shaping both siblings' behaviors.

The interpretation of sibling correlations depends critically on how the unobserved component, α, is treated. The fixed effects model controls for unobserved time-invariant heterogeneity in panel data by removing time-invariant family characteristics, thereby mitigating bias from unobserved heterogeneity across households. This provides more reliable estimates of the effects of observed covariates. Still, it also entails an important limitation: because α is differenced out, the contribution of shared but unobserved common factors to sibling similarity cannot be separately identified. Consequently, the decomposition is restricted to parents' socioeconomic status and the causal effect of the older sibling.

By contrast, a correlated random effects (CRE) model allows explicit estimation of α, making it possible to recover all three components, which are parents' socioeconomic status, the causal influence of the older sibling, and the contribution of shared but unobserved common factors. This provides a more comprehensive account of the mechanisms underlying sibling similarity in risky behaviors and extends prior research that has largely emphasized observed variables.

The empirical framework employed here integrates the strengths of both approaches. The fixed effects estimates offer a conservative benchmark by ensuring robustness against time-invariant omitted variables, while the correlated random effects estimates broaden the analysis by quantifying the role of unobserved common factors. Taken together, these complementary strategies demonstrate that the magnitude of sibling correlations and the decomposition structure may vary with the assumptions imposed, underscoring the value of presenting both approaches in tandem.

3.4.2 Baseline Structural Decomposition of Sibling Correlations

The analyses reveal that the common factor, a shared, unobserved family component often excluded in prior studies, emerges as a key explanatory source of sibling correlations. Notably, in both approaches, the estimated effect of the older sibling remains the smallest contributor. For certain risky behaviors, particularly drinking and violence, the common factor has a dominant explanatory power. This finding in line with evidence that risky behaviors such as drinking and violence are readily observed

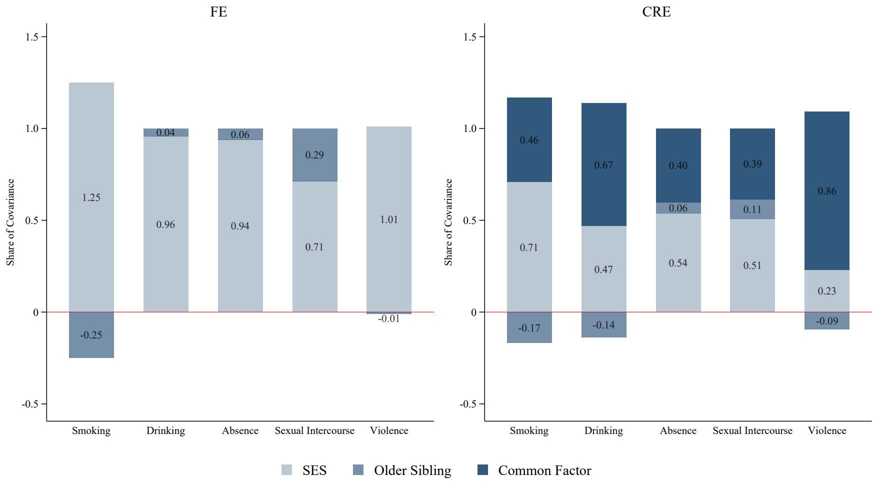

Fig. 6 Decomposition of Sibling Correlations in Risky Behavior (Full Sample)
Note: Bars show the decomposition of sibling correlations (ρ) into the contributions of parental socioeconomic status, older-sibling effect, and unobserved common factors. Results for the FE and CRE specifications are reported separately. Details on the decomposition method and calculation procedure appear in the Appendix.

within the household and become internalized through parental routines and disciplinary practices (Chatterji and Markowitz, 2001). It underscores the importance of accounting for the magnitude of the common factor's contribution, which may vary across types of risky behaviors.

Figure 6 illustrates a tendency for the contribution of the common factor to be excessively absorbed by parents' SES. This occurs because shared components differ in the fixed effects model, and their explanatory power is misallocated to SES. Three mechanisms help clarify this phenomenon. First, α affects both siblings simultaneously, differing in nature from the older sibling effect, which operates only through the elder child's behavior at a given time. For example, in the case of smoking, the CRE estimates assign a sizeable contribution to the common factor, 0.46, reflecting household-level rules, such as a no-smoking policy imposed on both children. Under the fixed effects model, however, this element is removed, inflating the explanatory power of SES to 1.25. The variation in α is thus more closely aligned with structural background variables like SES than the older sibling's causal influence. Second, parents' SES and α are empirically highly correlated. Parental education and economic resources are closely linked to household discipline, lifestyle, and parenting attitudes. When α is eliminated by differencing in the fixed effects model, its variation is naturally absorbed into the SES component. For instance, in school absence, parents' SES accounts for 0.54 and the common factor for 0.40 under the CRE, but in the fixed effects model, parents' SES rises to 0.94, which is almost the sum of the CRE estimates for parents' SES and the common factor. While parental education and income shape children's schooling, the fixed effects model estimates conflate these effects with broader unobserved aspects of the home environment. Third, by construction, the fixed effects model estimator redistributes the variance of time-invariant elements across the remaining observed covariates rather than eliminating it. Because parents' SES is the strongest candidate among the observed variables, it tends to inherit this residual variance, leading to an overstatement of its contribution. For example, in violence, the common factor is overwhelmingly large, 0.86, in the CRE, but once removed in the fixed effects model, the estimated contribution of parents' SES escalates to 1.01. The underlying source may lie in family conflict-resolution norms or the household's emotional climate. Yet, parents' SES appears to account for these cultural or environmental traits in the fixed effects model decomposition.

Taken together, these findings caution against interpreting SES contributions from a fixed effects model as exact measures of parents' background. When the common factor is not explicitly modeled, the fixed effects model can overstate the role of parents' SES, potentially obscuring the broader sources of sibling similarity in risky behaviors.

3.4.3 Structural Decomposition of Sibling Correlations by Gender Composition

In the analysis based on the full sample, the fixed effects model does not fully account for part of the shared environment and is instead attributed to parents' SES. As a result, sibling correlations appear to be excessively explained by parents' SES. Decompositions by sibling gender composition in Figure 7 and 8 reveal a similar redistribution pattern, while also uncovering clear differences in the direction and magnitude of the older sibling effect depending on the gender pairing.

For younger brothers, the fixed effects model shows a striking contrast by the gender of the older sibling. When the older sibling is a brother, the estimated effect is positive for most behaviors, except violence. By contrast, when the older sibling is a sister, a negative deterrent effect is observed for many behaviors, except smoking and violence. Under the CRE model, however, the contributions of older brothers and older sisters, apart from smoking, become similar in both magnitude and sign. That is, younger brothers display pronounced gender differences at the fixed effects stage, but these differences largely dissipate once the CRE model is applied. This suggests that a substantial portion of the effects estimated as parents' SES and the older sibling effect in the fixed effects model are, in fact, generated by the common factor, which is reallocated appropriately when the CRE model is used.

Differences between the fixed effects model and the CRE model estimates are also evident for younger sisters. In the fixed effects model, having an older sister is associated with a negative deterrent effect on absenteeism, whereas having an older brother is linked to positive effects for certain behaviors. In the CRE model, however, no clear pattern emerges from the older sibling's gender. The estimated effects appear irregular and lack a consistent direction. This indicates that, in the fixed effects model, the influence of the common factor is largely absorbed by parents' SES rather than being reflected in the estimated older sibling effect.

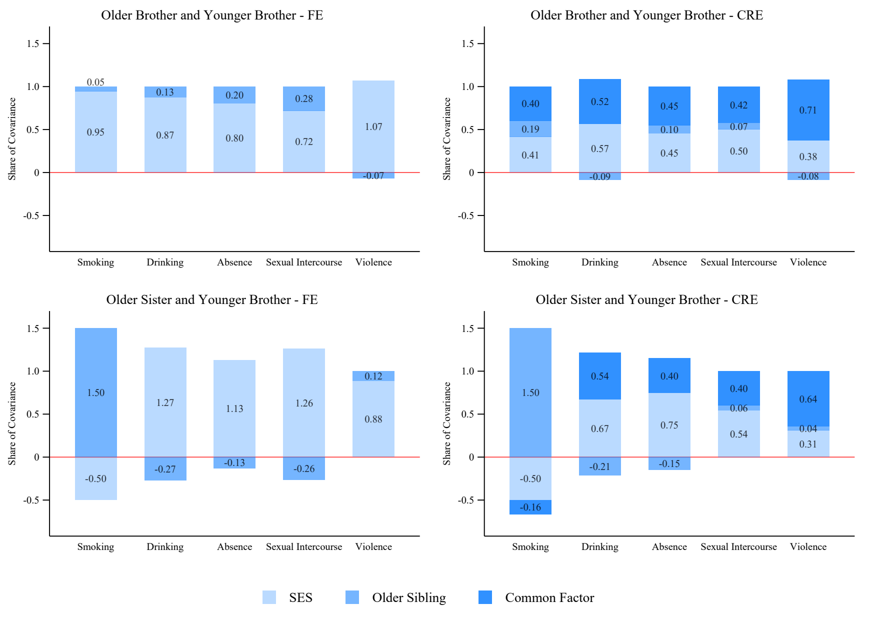

Fig. 7 Decomposition of Sibling Correlations in Risky Behavior by Younger Sibling's Gender: Boy
Note: Bars show the decomposition of sibling correlations (ρ) into the contributions of parental socioeconomic status, older-sibling effect, and unobserved common factors. Results for the FE and CRE specifications are reported separately. Details on the decomposition method and calculation procedure appear in the Appendix.

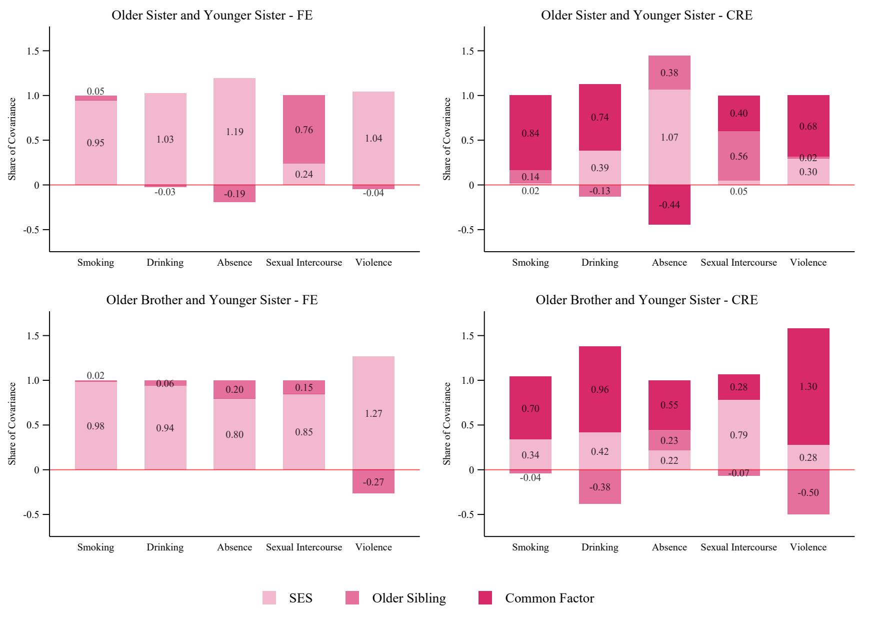

Fig. 8 Decomposition of Sibling Correlations in Risky Behavior by Younger Sibling's Gender: Girl
Note: Bars show the decomposition of sibling correlations (ρ) into the contributions of parental socioeconomic status, older-sibling effect, and unobserved common factors. Results for the FE and CRE specifications are reported separately. Details on the decomposition method and calculation procedure appear in the Appendix.

3.4.4 Structural Decomposition of Sibling Correlations by Age Gap

As Figure9 shows, the structural decomposition by age gap between siblings shows a distinct contrast between the fixed effects model and the CRE model. Under the fixed effects model, parents' SES exerts a dominant explanatory power, whereas the CRE model reveals a different pattern. In the group with a narrow age gap, parents' SES continues to retain explanatory strength, while the older sibling effect persists for certain behaviors. By contrast, in the group with a wider age gap, the explanatory power of parents' SES declines, and unobserved shared environmental components associated with parents' SES, captured by the common factor, become the primary source of variation for most behaviors except smoking.

Moreover, in the narrow gap group, the estimated sibling effect remains partly intact when moving from the fixed-effects specification to the CRE model, although its magnitude and direction shift somewhat. By contrast, in the wider gap group, the older sibling effect almost disappears for most behaviors, except smoking, once the CRE specification is employed. These differences reflect how parents' SES and shared family environments operate across developmental stages. When the age gap is small, parents' allocation of resources and educational practices are likely to be applied similarly to both children, allowing a sibling effect to emerge. When the age gap is large, however, differences in developmental timing may lead parents to allocate resources and enforce discipline differentially. In this case, sibling similarity is more plausibly explained by parents' SES and shared family norms, such as household discipline or atmosphere, rather than by the older sibling's direct influence.

4 Conclusion

Siblings raised in the same household under the same parents are likely to exhibit similar outcomes across a broad set. Understanding sibling similarities provides important clues for interpreting and predicting both short- and long-term child outcomes as well as inter-generational trajectories. However, a longstanding identification challenge has been distinguishing whether such resemblance stems from direct behavioral transmission between siblings or from the influence of shared family environments, which has significantly limited identification.

This paper investigates the structural sources of sibling similarity in adolescent risky behaviors and identifies the extent to which such resemblance reflects causal rather than merely correlational. By drawing on the first South Korean panel dataset investigating siblings, this study decomposes the observed similarity into the contributions of parental socioeconomic status (SES), spillover effects from older siblings, and unobserved common factors such as parenting practices and household norms.

I find that sibling similarity in risky behaviors cannot be explained solely by parental SES or a single sibling effect. I demonstrate this consistently using four complementary identification strategies. SES accounts for only a limited share of the overall correlation and sometimes functions as a suppressor that obscures or contaminates the underlying mechanisms within families. A substantial portion of the similarity instead stems from unobserved common factors, and these effects are particularly

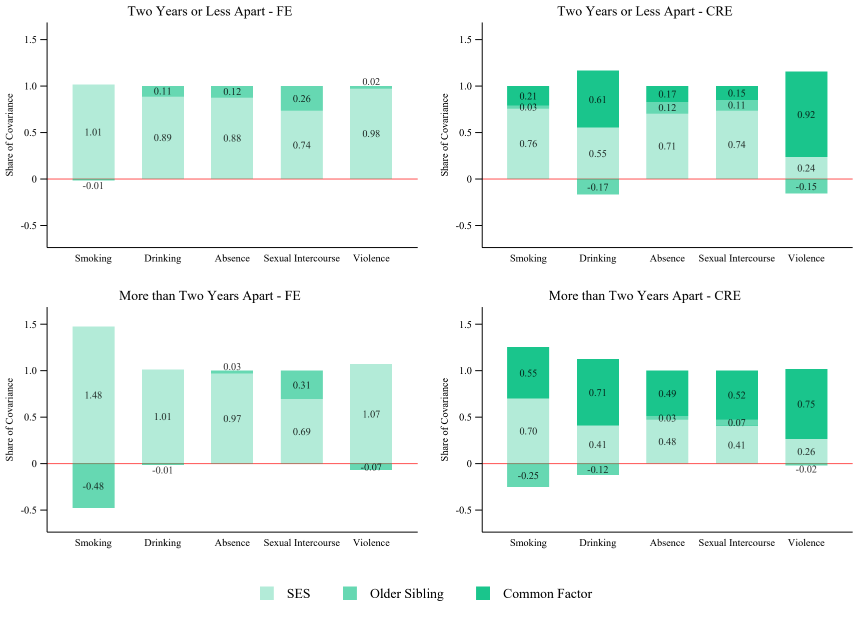

Fig. 9 Decomposition of Sibling Correlations in Risky Behavior by Age Gap
Note: Bars show the decomposition of sibling correlations (ρ) into the contributions of parental socioeconomic status, older-sibling effect, and unobserved common factors. Results for the FE and CRE specifications are reported separately. Details on the decomposition method and calculation procedure appear in the Appendix.

pronounced for behaviors like drinking and violence, where visibility and parental sanctioning are high. Once unobserved common factors are controlled, the primarily contemporaneous causal spillovers from older siblings weaken considerably or even shift in a deterrent direction, suggesting that normative learning or risk-avoidance mechanisms operate rather than simple imitation. The structural decomposition further shows that the direct effect of older siblings is relatively small. At the same time, SES and unobserved common factors explain most of the observed similarity, thereby providing stronger corroboration for the interpretations derived from the mechanism-based identification strategies. These patterns are especially salient among brother pairs and siblings with small age gaps. Across all specifications, it becomes evident that resemblance in risky behaviors is structured by socioeconomic background and family norms rather than simple peer transmission.

Nevertheless, two limitations remain. First, the analysis does not control for parental lifestyle habits or parenting styles, such as parents' smoking, drinking, or disciplinary practices. Since risky behavior often correlates with parental lifestyle, omitting these factors may bias estimates of common family effects. Second, the sample consists of adolescents from a low-fertility cohort, meaning families with multiple children are relatively limited. As a result, subgroup analyses by sibling gender composition or age gap may suffer from selection bias due to small cell sizes.

Despite these limitations, this study stands in contrast to existing research by demonstrating that neither a single external factor alone is sufficient to explain adolescent risky behavior. In particular, the analytical mechanism incorporating the importance of unobserved factors, such as parenting styles and parental attitudes, contributes to a more accurate understanding of the nature of adolescent risky behaviors. Moreover, this approach holds the potential to be applied not only to adolescents but also to empirical research on peer effects across different age groups in the South Korean context, where studies on behavioral similarity remain limited. By empirically analyzing sibling correlations in the behaviors of contemporary Korean adolescents and identifying the causal influences embedded within such similarity, this study carries substantial social significance as the first of its kind in South Korea and establishes a necessary frame of reference for interpreting adolescent behavior within the family context.

5 Appendix

Appendix A. Additional Figures

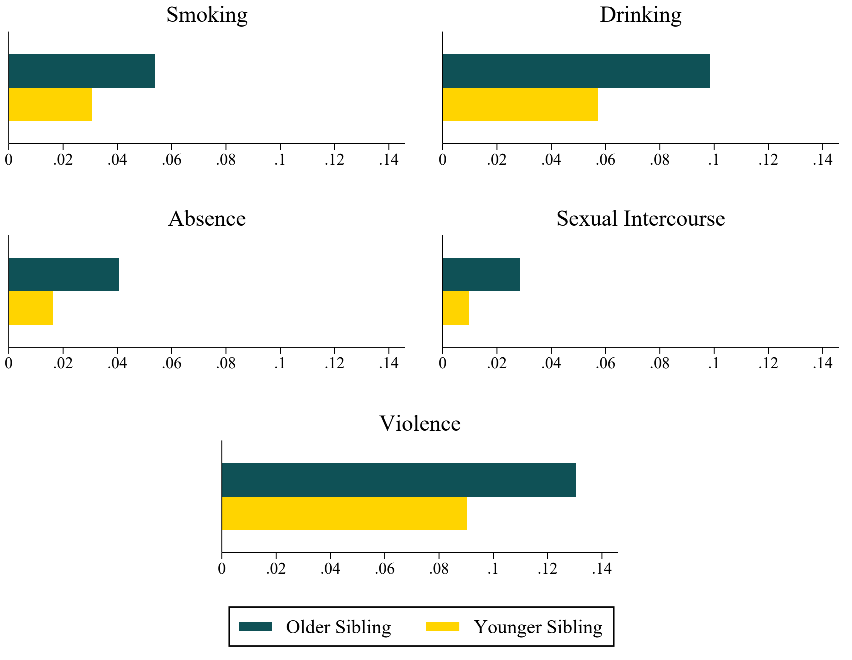

Fig. A1 Average Risky Behavior of Older and Younger Siblings
Notes. Bars display the mean prevalence of each risky behavior for older and younger siblings in the sample. All variables are coded as binary indicators equal to 1 if the behavior was ever reported.

Appendix B. Estimation Strategy: Fixed Effects via Frisch-Waugh-Lovell

Estimating γ requires addressing that siblings share the same family environment, including unobserved parental characteristics and household-level conditions. If these unobserved components are correlated with both R^o and R^y, then regressing R^y on R^o would conflate causal spillovers with spurious similarity from shared background factors. This appendix presents the fixed effects (FE) estimator based on the Frisch–Waugh–Lovell (FWL) theorem, which differences out all between-household variation and identifies γ using within-household contrasts. The starting point is

$$R^y = \gamma R^o + \mathbf{X}\theta + \alpha + u^y \tag{13}$$

where α captures all household-level unobserved components. Identification based on the assumption $\mathrm{Cov}(u^y, R^o) = 0$ after differencing out α through the FE transformation. Let H be the $n \times G$ matrix of household dummies. Then,

$$\mathscr{P}_H \equiv H(H'H)^{-1}H' \quad \text{and} \quad \mathscr{M}_H \equiv I - \mathscr{P}_H$$

which yield the household mean and the within-household deviation, respectively. By the FWL theorem, OLS with household dummies is equivalent to OLS applied to transformed variables

$$\tilde{Z} \equiv \mathscr{M}_H Z, \quad Z \in \{R^y, R^o, \mathbf{X}\}$$

The fixed effect estimating equation becomes

$$\tilde{R}^y = \gamma^{FE}\tilde{R}^o + \tilde{\mathbf{X}}\theta^{FE} + \tilde{u}^y \tag{14}$$

which relies exclusively on within-household variation. When the focus is on γ, the FWL theorem implies

$$\widehat{\gamma}^{FE} = \frac{\tilde{R}^{o\prime}\mathscr{M}_{\tilde{\mathbf{X}}}\tilde{R}^y}{\tilde{R}^{o\prime}\mathscr{M}_{\tilde{\mathbf{X}}}\tilde{R}^o}, \qquad \mathscr{M}_{\tilde{\mathbf{X}}} \equiv I - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'$$

which corresponds to regressing the residual of $\tilde{R}^y$ on the residual of $\tilde{R}^o$ after partialling out $\tilde{\mathbf{X}}$.

The fixed effects estimator can be implemented in three steps:

1. **Within-household residualization:**

$$\tilde{R}^y = \mathscr{M}_H R^y, \quad \tilde{R}^o = \mathscr{M}_H R^o, \quad \tilde{\mathbf{X}} = \mathscr{M}_H \mathbf{X}$$

2. **Partialling out observed SES covariates:** Regress $\tilde{R}^y$ on $\tilde{\mathbf{X}}$ and store the residual e^y. Regress $\tilde{R}^o$ on $\tilde{\mathbf{X}}$ and store the residual e^o.

3. **Estimation of γ:** Regress e^y on e^o using OLS to obtain $\widehat{\gamma}^{FE}$. Standard errors are clustered at the household level.

When exactly two siblings are observed per household, $\mathscr{M}_H$ reduces to demeaning by the household average. In this case, (14) coincides with the familiar within transformation and remains algebraically equivalent to OLS with household fixed effects.

Appendix C. Structural Decomposition of Sibling Correlation

To decompose the observed sibling correlation into additive structural sources, start from the linear equation:

$$R^y = \gamma R^o + \mathbf{X}\theta + \alpha + u^y \tag{15}$$

Taking the covariance with R^o yields the raw co-movement:

$$\begin{align} \mathrm{Cov}(R^y, R^o) &= \mathrm{Cov}(\gamma R^o + \mathbf{X}\theta + \alpha + u^y, R^o) \tag{16} \\ &= \gamma \mathrm{Var}(R^o) + \mathrm{Cov}(\mathbf{X}\theta, R^o) + \mathrm{Cov}(\alpha, R^o) + \mathrm{Cov}(u^y, R^o) \tag{17} \end{align}$$

Imposing orthogonality of the idiosyncratic innovation to the older sibling's outcome eliminates the final term:

$$\mathrm{Cov}(R^y, R^o) = \gamma \mathrm{Var}(R^o) + \mathrm{Cov}(\mathbf{X}\theta, R^o) + \mathrm{Cov}(\alpha, R^o) \tag{18}$$

Each component can be rewritten using the mean-deviation identity $\mathrm{Cov}(A,B) = E[(A - E[A])(B - E[B])]$:

$$\begin{align} \mathrm{Cov}(\mathbf{X}\theta, R^o) &= E[(\mathbf{X}\theta - E[\mathbf{X}\theta])(R^o - E[R^o])], \tag{19} \\ \mathrm{Cov}(\alpha, R^o) &= E[(\alpha - E[\alpha])(R^o - E[R^o])], \tag{20} \\ \mathrm{Var}(R^o) &= E\left[(R^o - E[R^o])^2\right] \tag{21} \end{align}$$

Normalizing by the marginal standard deviations defines the correlation coefficient:

$$\rho \ = \ \frac{\mathrm{Cov}(R^y, R^o)}{\sqrt{\mathrm{Var}(R^y)\,\mathrm{Var}(R^o)}} \tag{22}$$

Using the covariance expressions above, the correlation coefficient can be written as the sum of three components:

$$\rho = \underbrace{\frac{\gamma \mathrm{Var}(R^o)}{\sqrt{\mathrm{Var}(R^y)\,\mathrm{Var}(R^o)}}}_{\text{Effect of the older sibling}} + \underbrace{\frac{\mathrm{Cov}(\mathbf{X}\theta, R^o)}{\sqrt{\mathrm{Var}(R^y)\,\mathrm{Var}(R^o)}}}_{\text{Contribution of parents' SES}} + \underbrace{\frac{\mathrm{Cov}(\alpha, R^o)}{\sqrt{\mathrm{Var}(R^y)\,\mathrm{Var}(R^o)}}}_{\text{Effect of unobserved common factors}} \tag{23}$$

References

Abboud, T., Bellou, A., and Lewis, J. (2024). The long-run impacts of adolescent drinking: Evidence from zero tolerance laws. *Journal of Public Economics*, 231:105066.

Aizer, A. (2004). Home alone: Supervision after school and child behavior. *Journal of public economics*, 88(9-10):1835–1848.

Aizer, A., Eli, S., Ferrie, J., and Lleras-Muney, A. (2016). The long-run impact of cash transfers to poor families. *American Economic Review*, 106(4):935–971.

Altonji, J. G., Cattan, S., and Ware, I. (2017). Identifying sibling influence on teenage substance use. *Journal of Human Resources*, 52(1):1–47.

Argys, L. M., Rees, D. I., Averett, S. L., and Witoonchart, B. (2006). Birth order and risky adolescent behavior. *Economic Inquiry*, 44(2):215–233.

Averett, S. L., Argys, L. M., and Rees, D. I. (2011). Older siblings and adolescent risky behavior: does parenting play a role? *Journal of Population Economics*, 24(3):957–978.

Balsa, A., Gandelman, N., and Roldán, F. (2018). Peer and parental influence in academic performance and alcohol use. *Labour Economics*, 55:41–55.

Balsa, A. I., Giuliano, L. M., and French, M. T. (2011). The effects of alcohol use on academic achievement in high school. *Economics of education review*, 30(1):1–15.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., Roth, J., and Simonsen, M. (2021). Sibling spillovers. *The Economic Journal*, 131(633):101–128.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under pressure? the effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1):119–153.

Booth, A. L. and Kee, H. J. (2009). Birth order matters: the effect of family size and birth order on educational attainment. *Journal of population economics*, 22(2):367–397.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1):211–228.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American economic review*, 92(5):1308–1334.

Chatterji, P. and Markowitz, S. (2001). The impact of maternal alcohol and illicit drug use on children's behavior problems: evidence from the children of the national longitudinal survey of youth. *Journal of health economics*, 20(5):703–731.

Conley, D. and Glauber, R. (2005). Sibling similarity and difference in socioeconomic status: life course and family resource effects.

Cook, P. J. and Moore, M. J. (1993). Drinking and schooling. *Journal of health economics*, 12(4):411–429.

Cools, A. and Patacchini, E. (2017). Sibling gender composition and women's wages. Technical report, IZA Discussion Papers.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *JoUrnal of economIc lIteratUre*, 47(1):87–122.

Currie, J. and Stabile, M. (2003). Socioeconomic status and child health: why is the relationship stronger for older children? *American Economic Review*, 93(5):1813–1823.

Dahl, G. B., Rooth, D.-O., and Stenberg, A. (2024). Intergenerational and sibling spillovers in high school majors. *American Economic Journal: Economic Policy*, 16(3):133–173.

Eriksson, K. H., Hjalmarsson, R., Lindquist, M. J., and Sandberg, A. (2016). The importance of family background and neighborhood effects as determinants of crime. *Journal of Population Economics*, 29:219–262.

Fletcher, J., Hair, N. L., and Wolfe, B. L. (2012). Am i my brother's keeper? sibling spillover effects: The case of developmental disabilities and externalizing behavior. Technical report, National Bureau of Economic Research.

Grossman, M. (2005). Individual behaviours and substance use: the role of price. In *Substance use: Individual behaviour, social interactions, markets and politics*, pages 15–39. Emerald Group Publishing Limited.

Gruber, J. (2009). *Risky behavior among youths: An economic analysis*. University of Chicago Press.

Harris, J. E. and López-Valcárcel, B. G. (2008). Asymmetric peer effects in the analysis of cigarette smoking among young people in the united states, 1992–1999. *Journal of health economics*, 27(2):249–264.

Kim, H. W., Kim, Y. K., Paik, J. Y., Hong, C. B., Lee, K., Park, T.-J., and Kim, J. (2017). The relationship between smoking and suicidal behavior in korean adolescents: 12th korea youth risk behavior web-based survey. *Korean Journal of Health Promotion*, 17(4):219–233.

Kim, Y., Choi, S., Chun, C., Park, S., Khang, Y.-H., and Oh, K. (2016). Data resource profile: the korea youth risk behavior web-based survey (kyrbs). *International journal of epidemiology*, 45(4):1076–1076e.

Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, 120(1):87–130.

Koch, S. F. and Ribar, D. C. (2001). A siblings analysis of the effects of alcohol consumption onset on educational attainment. *Contemporary Economic Policy*, 19(2):162–174.

Korea Disease Control and Prevention Agency (2021). Korea youth risk behavior survey fact sheet. Retrieved from `https://www.kdca.go.kr/board/board.es?mid=a20501010000&bid=0015&list_no=368345&cg_code=&act=view&nPage=6&newsField=202009`.

Lee, S.-h. and Hoe, M. (2018). A structural analysis of the relationship among depression, aggressiveness, social withdrawal, and school adjustment in adolescents: Focusing on differences in delinquency experiences. *Health and Social Welfare Review*, 38(4):251–289.

Louis, F. R. B. S. (2024). Federal reserve economic data (fred). `https://fred.stlouisfed.org/series/AEXKOUS`. Accessed: 2025-09-30.

Lundborg, P. (2006). Having the wrong friends? peer effects in adolescent substance use. *Journal of health economics*, 25(2):214–233.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Mazumder, B. (2008). Sibling similarities and economic inequality in the us. *Journal of Population Economics*, 21(3):685–701.

Mechoulan, S. and Wolff, F.-C. (2015). Intra-household allocation of family resources and birth order: evidence from france using siblings data. *Journal of Population Economics*, 28(4):937–964.

Nicoletti, C. and Rabe, B. (2019). Sibling spillover effects in school achievement. *Journal of Applied Econometrics*, 34(4):482–501.

Oettinger, G. S. (2000). Sibling similarity in high school graduation outcomes: Causal interdependency or unobserved heterogeneity? *Southern economic journal*, 66(3):631–648.

Rodgers, J. L. and Rowe, D. C. (1988). Influence of siblings on adolescent sexual behavior. *Developmental Psychology*, 24(5):722.

Solon, G., Corcoran, M., Gordon, R., and Laren, D. (1991). A longitudinal analysis of sibling correlations in economic status. *Journal of Human resources*, pages 509–534.

Tippett, N. and Wolke, D. (2015). Aggression between siblings: Associations with the home environment and peer bullying. *Aggressive behavior*, 41(1):14–24.

Valencia, M. L. C., Tran, B. T., Lim, M. K., Choi, K. S., and Oh, J.-K. (2019). Association between socioeconomic status and early initiation of smoking, alcohol drinking, and sexual behavior among korean adolescents. *Asia Pacific Journal of Public Health*, 31(5):443–453.

World Health Organization (2024). Adolescent and young adult health. Retrieved from `https://www.kdca.go.kr/board/board.es?mid=a20501010000&bid=0015&list_no=368345&cg_code=&act=view&nPage=6&newsField=202009`.