Appendix 5. Evidence summaries RQ1

Question (PICO)	In children and adolescents aged 5 to 18 years (P), what risk or protective factors related to how they interact/communicate with their peers (I) are associated with their mental health (O)?
Search Strategy	See Appendix 1
Search date	11/5/2021
In/Exclusion criteria	See Appendix 2

Content

Thematic category: (Pro)social behaviour

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Bowker, 2014, USA	Observational: prospective cohort study (2 timepoints in 6 month period)	264 students (age range not available, mean age 12.08 years at T1, SD not available), from two middle schools (grade 6)	Being the recipient of prosocial peer treatment: peer nomination items were used to assess prosocial peer treatment, using the statement 'Someone who you are nice to'. Participants were allowed to make unlimited nominations from their grade and school (self-reported). Nominations received by each participant were summed, proportionalized and standardized within grade/school.	Outcomes measured: - depressive symptoms: assessed with the short (10-item) form of the Child Depression Inventory (CDI) - anxiety: peer nomination items were used to assess anxious-withdrawal (4-items, e.g. 'Someone who is very shy') Statistics: hierarchical linear regression analyses, with five models (two for concurrent analyses and three

				for longitudinal analyses)
Toseeb, 2020, UK	Observational: retrospective cohort study (2 timepoints with 5 year interval)	13287 children (age range not available, mean age 7 years (SD unknown) at T1 and 11 years (SD unknown) at T2)	Prosocial behaviour: assessed by the parent-reported Prosocial subscale of the parent-reported Strength and Difficulties Questionnaire and administered when the child was 7 years old (more details see below)	Outcomes measured at 11 years old - bullying: the child was asked " how often do you hurt or pick on your brothers or sisters on purpose?"; perpetrated at least once a week but not victimized Statistics: single multinomial logistic regression model; all of the significant predictors from the previous set of models were entered as predictors

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
Depressive symptoms (T1 --> T2)	Prosocial peer treatment	Not statistically significant: £† (p>0.05)	1, 264 §	Bowker, 2014
Anxiety (T1 --> T2)		Not statistically significant: £† (p>0.05)		
Bullying (T1 --> T2)	Prosocial behaviour	Not statistically significant: £ RR: 0.89, 95%CI [0.80;1.00] (0.01 < p < 0.05; p in paper set at 0.01)	1, 13287	Toseeb, 2020

RR: risk ratio, CI: confidence interval
£ No raw data/SD's available
† Imprecision (lack of data)
§ Imprecision (limited sample size)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Bowker, 2014	Yes, no random sampling, 70% consent rate of 2 participating middle schools (unclear on what basis schools were selected)	Yes, exposure and outcomes were self-reported.	No, in all models, acceptance was entered at Step 1 as a covariate and in the longitudinal analyses, baseline outcomes were also entered at Step 1.	Yes, it is important to note that several participants did not complete the depression measure (127 participants missing at T1 and 98 participants missing at T2) or mutual friendship measure (69 missing students). However, statistical analysis did not reveal any differences on the	Yes, no transparency for model fit and associated non-statistically significant results

				peer nomination variables between those students who completed the depression or mutual friendship measures and those who did not.	
Toseeb, 2020	No, the sample is representative of the UK population and was drawn from the entire live birth cohort of the UK between the years 2000-2001.	Yes, sibling bullying was self-reported. Exposure was parent-reported.	No, the extensiveness of data collected from the MCS families enabled the inclusion of a number of covariates in all of the statistical models.	Yes, some variables had high levels of missing data. Whilst 50 imputed datasets were generated to deal with missing data, this should be borne in mind when interpreting the findings.	No, no indication

Certainty of the body of evidence

Prosocial peer treatment	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample size/Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Prosocial behaviour	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Conclusion	*Prosocial peer treatment* There is limited evidence concerning the risk of depression or anxiety in case of being the recipient prosocial peer treatment (i.e. peers are nice). A statistically significant decreased risk of depression or anxiety in case of being the recipient of prosocial peer treatment could not be demonstrated (Bowker 2014). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. *Prosocial behaviour* There is limited evidence showing no correlation between previous prosocial behavior and the risk of bullying at later age. It was shown that prosocial behavior at 7 years did not result in a statistically significant decreased risk of bullying at 11 years (Toseeb 2020). Evidence is of very low certainty.
Reference(s)	**Articles**

| | Bowker JC. *Prosocial peer treatment and the psychosocial outcomes associated with anxious-withdrawal.* Inf Child Dev 2014, 23:314-322 |
| | Toseeb U, McChesney G, Dantchev S, Wolke D. *Precursors of sibling bullying in middle childhood: Evidence from a UK-based longitudinal cohort study.* Child Abuse Negl 2020, 108:104633 |

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Prosocial behaviour	Toseeb, 2020	Prosocial subscale of the Strengths and Difficulties Questionnaire (SDQ)	Contains 5 items: "[child] is helpful if someone is hurt, upset or feeling ill"; "[child] is considerate of other people's feelings"; "[child] shares readily with other children, for example toys, treats, pencils"; "[child] is kind to younger children"; "[child] often volunteers to help others (parents, teachers, other children)" Responses were coded on a three-point scale (0=not true, 1 = somewhat true, 2 = certainly true). There were 5 items with a total score ranging from 0 to 10 (higher scores indicate higher levels of prosociality).

Thematic category: Factors related to the content of communication

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Bastin, 2014, Belgium	Observational: prospective cohort study (2 timepoints with 3-month interval)	371 pupils (5th and 7th grade), aged 9.42-15.00 years (mean age 11.73±1.10 years), attending 9 schools in Flanders; 357 pupils at follow-up	Co-brooding and co-reflection with closest same-sex friend (selected items from the Co-Rumination Questionnaire (CRQ))	Identified from reference list from 'Bastin 2021' Outcomes measured: depressive symptoms during past 2 weeks, assessed by Children's Depression Inventory Statistics: hierarchical regression analysis
Bastin, 2015, Belgium	Observational: prospective cohort study (4 timepoints in a 1 year period)	368 pupils (5th and 7th grade), aged 9.42-15.00 years (mean age 11.72±1.10 years), attending 9 schools in Flanders; 291 who completed all assessments.	Co-rumination with closest same-sex friend (27-item CRQ)	Identified from reference list from 'Bastin 2021' Outcomes measured: depressive symptoms during past two weeks, assessed by Children's

		[Possible overlap between this patient population and the one from Bastin 2014 (but longer follow-up)]		Depression Inventory Statistics: hierarchical linear multilevel modelling
Bastin, 2018a, Belgium	Observational: prospective cohort study (2 timepoints with 1-year interval)	665 adolescents from the 7th grade, aged 11.3-14.9 years (mean age 12.7±0.4 years), attending 7 secondary schools in Flanders; 543 adolescents at T2, of those, 70 adolescents participated for the first time.	Co-dampening and Co-enhancing with same-sex best friend: two types of interpersonal sharing of positive feelings between peers (Co-Dampening and Co-Enhancing Questionnaire (CoDEQ))	Identified from reference list from 'Bastin 2021' Outcomes measured: - depressive symptoms during past 2 weeks assessed by Children's Depression Inventory - anhedonic symptoms: Leuven Anhedonia Self-report Scale Statistics: cross-lagged analyses using structural equation modelling. Post-hoc subgroup analyses were available for adolescents with a stable best friend (i.e. same friend at T1 as T2)
Bastin, 2018b, Belgium	Observational: prospective cohort study (2 timepoints with 2-month interval)	Out of 465 pupils from 5th to 10th grade from 11 schools in Flanders, only participants naming the same person as their best friend at both assessments were included in the analyses: 313 adolescents, aged 9-17 years (mean age 13±2.07 years)	Co-brooding and co-reflection with closest same-sex friend (selected items from the CRQ)	Identified from reference list from 'Bastin 2021' Outcomes measured: depressive symptoms during past 2 weeks assessed by Children's Depression Inventory Statistics: structural equation modelling. Subgroup analyses were done for both boys and girls
Bastin, 2021, Belgium	Observational: prospective cohort study (3 timepoints with 1 year-interval)	1549 adolescents, aged 9-17 years (mean age 12.93±1.46 years), from 5th to 9th grade in seven schools in Flanders at T1; 1188 individuals at T2, 899 at T3.	Co-rumination with closest, same-sex friend (9-item version of CRQ) Research questions were also examined using the co-brooding component of co-rumination. For this, 6 items of the original CRQ were used, corresponding with the content areas speculation about	Outcomes measured: - depressive symptoms during past two weeks assessed by Children's Depression Inventory (self-report) - brooding rumination: self-

			consequences and focusing on negative feelings.	report rumination subscale of the Children's Response Styles Questionnaire Statistics: mediation analyses and cross-lagged path analyses
Blodgett Salafia, 2010, USA	Observational: prospective cohort study (4 timepoints with 1 year intervals)	85 adolescent girls (mean age at T1 10.59±0.52 years, range 10-12 years) from 5th grade (at T1) in a primary school in a Midwestern city	Peer discussion about dieting, assessed by a single item: "How often have your friends talked about wanting to lose weight?"	Identified from the systematic review by Webb 2013 Outcome measured: body dissatisfaction at T2 (Body Dissatisfaction subscale of the Eating Disorders Inventory, EDI) Statistics: structural equation modelling (SEM) [Data from timepoints 3 and 4 (7th and 8th grade) were not extracted because no direct links with the RF at T1 were reported]
Borschmann, 2020, Australia	Observational: prospective cohort study (4 timepoints with a 1-year interval)	1239 participants, aged 8-9 years at T1, 11-12 years at T4 (mean age and SD unknown) from 43 primary schools in Melbourne.	Overt victimisation (teasing) at T1, T2 and T3, assessed with selected items from Gatehouse Bullying Scale	Outcomes measured at T4: Self-harm: assessed by asking "In the past 12 months have you ever hurt yourself on purpose or done anything that might have harmed you or even killed you?" If participants responded "yes", they were then asked to describe what they did. Self-harm was defined as any behaviour fitting into one of five categories: (1) cutting/burning, (2) self-poisoning, (3) self-battering, (4) non-recreational risk-taking (e.g., reckless behaviour near traffic), or (5) other self-harm. Statistics: Multivariate model.
Dirghangi, 2015, Canada	Observational: prospective cohort study	364 participants (aged 12 years, mean age unknown)	Co-rumination with best friend (4-item abbreviated version of the CRQ)	Identified from reference list from 'Bastin 2021'

	(2 timepoints with 1-year interval)	belonging to 182 monozygotic (MZ) twin pairs born between 1995 and 1998 in the greater Montreal area.		Outcomes measured: - depressive symptoms during past two weeks assessed by Children's Depression Inventory (CDI) - anxiety during the previous month: Children's Manifest Anxiety Scale Statistics: path analysis using monozygotic twin difference score, or using individual raw scores (one member of each twin dyad was randomly selected)
Ellickson, 2004, USA	Observational: prospective cohort study (control arm of an experimental study; 4 timepoints with 1-year interval)	1955 students (mean age and age range unknown) in 7th grade who were participating in 'Project ALERT', a drug prevention program for middle-school students. At subsequent waves, the number of individuals who completed the assessment and who did not have missing information on marijuana use was 1713 at T2 (8th grade), 1590 at T3 (9th grade), and 1557 at T4 (10th grade)	Friend approval of marijuana use: rated in terms of how friends would feel if they found out that the participant used marijuana (1 = they would disapprove and stop being my friends to 4 = they would approve).	Outcomes measured: marijuana use, confirmed by saliva sample Statistics: multivariate logistic regression
Hankin, 2010, USA	Observational: prospective cohort study (4 timepoints with 5-week intervals)	350 students, aged 11-17 years (mean age 14.5±1.40 years at baseline), from 6th to 10th grades from 5 Chicago area schools; 303 students at T2, 308 at T3, 345 at T4	Co-rumination with same-sex friend at T1, T2 and T4 (9-item abbreviated version of the CRQ)	Identified from reference list from 'Bastin 2021' Outcomes measured: - depressive symptoms assessed by Children's Depression Inventory - anhedonic depression: selected items from the Children's Depression Inventory (CDI) - anxiety: Mood and Anxiety Symptom

				Questionnaire (MASQ) - externalizing problems assessed by Strengths and Difficulties Questionnaire (SDQ) (conduct factor) - internalizing problems assessed by Strengths and Difficulties Questionnaire (SDQ) (emotional factor) Statistics: structural equation model and hierarchical linear modelling [for hierarchical linear modelling only slopes were extracted, not intercepts; data from model investigating mediating role of stressors were not extracted]
Helfert, 2011, Germany	Observational: prospective cohort study (2 timepoints with 1-year interval)	236 girls, aged 11-15 years (mean age 13.98±0.89 years) and 193 boys, aged 12-16 years (mean age 14.13±0.97 years) from grades 7-9 in three German high schools	Peer teasing, assessed with a subscale of the Appearance-related Social Pressure Questionnaire (FASD)	Identified from the systematic review by Webb 2013 Outcomes measured: Weight concerns, assessed with the Body Dissatisfaction subscale of the German version of the revised Eating Disorder Inventory (EDI-2). Statistics: hierarchical multiple regression analyses
Jones, 2004, USA	Observational: prospective cohort study (2 timepoints with 1-year interval)	158 adolescents (mean age 12.5 year, SD unknown, at T1) from 2 middle schools and 146 adolescents (mean age 15.5 years, SD unknown, at T1) from 2 high schools in a metropolitan district of a major northwestern city	- Appearance conversations: 5 items assessed the frequency with which students reported talking about their bodies and appearance enhancements with their friends. E.g. "My friends and I talk about what we can do to look our best" and "My friends and I talk about what we would like our bodies to look like." - Appearance teasing: 4 items assessed teasing by	Identified from reference list of Webb 2017 Outcomes measured: Body image dissatisfaction: assessed with the Body Dissatisfaction subscale from the Eating Disorder Inventory Statistics: multivariate

			friends: 2 items each about teasing from "girls who are my friends" and "boys who are my friends." The teasing by other peers was indicated by responses to items about "girls who I don't hang out with" and "boys who I don't hang out with." E.g. "Boys who are my friends tease me or make fun of me because of my body" and "Girls who I don't hang out with tease me or make fun of the size or shape of my body." Responses to each item ranged from 1 (never) to 5 (very often).	multiple regression models
Kam, 2013, USA	Observational: prospective cohort study (4 rounds: T1: Nov 1999 to June 2001; T2: July 2001 to June 2002; T3: July 2002 to June 2003; T4: July 2003 to June 2004)	Data were collected through the National Survey of Parents and Youth (NSPY) to evaluate the National Youth Anti-Drug Media Campaign. 2749 children and adolescents, aged 9-18 years, were included in the study. At T1, the mean age was 13.35±1.3 years and by T4 the mean age was 16.45±1.0 years.	Best-friend communication against drugs (at T2)	Outcomes measured: marijuana, inhalant, and ecstasy use Statistics: multiple mediation model
Kam, 2015, USA	Observational: prospective cohort study (3 timepoints with 3 to 4-month intervals)	277 Latino and 350 non-Latino white students in 6th–8th-grade-students from 3 rural Illinois public schools. Mean age was 12.4±1.0 years.	Best-friend communication against substance use (adapted from 6 items from the Targeted child Communication about Alcohol Scale)	Outcomes measured: marijuana use in past 3 months Statistics: multigroup structural equation modeling
Laird, 2013, USA	Observational: prospective cohort study (2 timepoints with 1-year interval) [study is part of a larger 3-wave longitudinal study]	181 adolescents in 6th and 7th grade from public schools in Southern USA, mean age 12.4±0.53 years at T1; N=148 at T2 (mean age 13.4±0.56 years)	Secrecy from friends (10-item Self-Concealment Scale in reference to secrecy from best friends)	Outcomes measured: - depression during the past month: 6-item Modified Depression Scale - antisocial behaviour during the past month: 26 items from the Problem Behavior Frequency Scale Statistics: polynomial regression equations
Latina, 2015, Italy	Observational: prospective	709 adolescents, aged 13 to 18 years	Co-rumination with one's best friend (9-item	Outcomes measured:

9

	cohort study (2 timepoints with 6-month interval)	(mean age 15.53±1.03 years), attending 3 different types of schools in Northern Italy	abbreviated version of the CRQ)	- self-harm: non-suicidal self-injury during the last year, assessed using a self-reported 6-item validated scale - Depressive symptoms during previous two weeks: Short Mood and Feeling Questionnaire Statistics: path analysis: maximum likelihood estimation with robust standard errors
Ohannessian, 2021, USA	Observational: prospective cohort study (3 timepoints with 6-month intervals)	1205 7th and 8th grade students, mean age 12.75±0.71 years, attending five middle schools in the southern New England region of the USA	Co-rumination with friends (at T2) assessed using the short version of CRQ: the CRQ-S	Outcomes measured: - depressive symptoms in past week: Center for Epidemiological Studies Depression Scale for Children - anxiety symptoms in last 3 months: Screen for Child Anxiety Related Disorders. Statistics: path analysis in which co-rumination was investigated as mediator between social media use and depression/anxiety
Ojeda, 2019, Spain	Observational: prospective cohort study (2 timepoints with 4-month interval, longitudinal study)	1736 Spanish High School students, 12-16 years (mean age 13.60±1.25 years), at T1; 1456 students at T2	Sexting: four types were distinguished: 1) sending, 2) receiving, 3) third-party forwarding and 4) receiving sexts via an intermediary [only data on sending/receiving from boyfriend/girlfriend was extracted, as it was not specified whether the third party was also a peer]	Outcomes measured: - cyberbullying in the last two months: aggression subscale of the European Cyberbullying Intervention Project Questionnaire, ECIPQ - bullying perpetration: the aggression subscale of the European Bullying Intervention Project Questionnaire was used. The cut-off point for both bullying measures was that the young person had aggressed against another "at least once or twice a

				month" in the last two months. Statistics: cross-lagged panel analyses
Paxton, 2006, Australia	Observational: prospective cohort study (2 timepoints with 5-year interval)	440 girls (mean age 12.7±0.76 years at T1) and 366 boys (mean age 12.8±0.76 years at T1) from 31 Minnesota (USA) schools	Weight-based teasing, assessed with 2 items measuring the frequency of being teased about weight and appearance on 5-point scales ranging from 1 (never) to 5 (at least once a week). Higher scores indicated more frequent teasing.	Identified from reference list from Webb 2017 Outcomes measured: Body dissatisfaction, assessed with a modified version of the Body Shape Satisfaction Scale. Statistics: Multivariable models
Rose, 2007, USA	Observational: prospective cohort study (2 timepoints with 6-month interval)	1048 students from 3rd, 5th, 7th and 9th grade in four Midwestern school districts at T1; 999 students at T2; complete data for 604 participants with a reciprocal friend. Mean age or age range unknown.	Co-rumination with reciprocal same-sex friends (27-item CRQ)	Outcomes measured: - depressive symptoms weeks assessed by Children's Depression Inventory - anxiety: Revised Children's Manifest Anxiety Scale Statistics: hierarchical regression analysis
Schwartz-Mette, 2016, USA	Observational: prospective cohort study (2 timepoints with 9-month interval)	480 adolescents (240 friendship dyads) in 7th (mean age 13.03±0.40 years) and 10th (mean age 16.04±0.45 years) grades from a public school district in a Midwestern university town, 333 adolescents at T2 Data for the current study were drawn from a larger project involving adolescents and their same-sex friends.	Co-rumination with reciprocal friends (27-item CRQ)	Outcomes measured: depression contagion between dyads: current (past week) depressive symptoms were measured using the Center for Epidemiological Studies Depression Scale Statistics: Actor-Partner Interdependence Model within a structural equation modelling framework
Starr, 2009, USA	Observational: prospective cohort study (2 timepoints with 1-year interval)	83 adolescent girls (mean age at T1 13.45±0.68 years), in 7th and 8th grade in 3 school districts in Suffolk County, New York	Co-rumination with friends (27-item CRQ)	Outcomes measured: - depressive symptoms: Child Depression Scale from the Center for Epidemiological Studies - social anxiety: Social Anxiety Scale for Adolescents

				Statistics: partial correlations between T1 co-rumination and T2 depressive symptoms, controlling for T1 depressive symptoms
Stone, 2011, USA	Observational: prospective cohort study (5 timepoints with 6-month intervals)	106 adolescents, aged 11-15 years (median 13 years), from a major midwestern city in the USA	Co-rumination with same-sex friend (9-item abbreviated version of the CRQ)	Outcomes measured: Depression onset, assessed with Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime Version Statistics: survival analysis
Stone, 2015, USA	Observational: prospective cohort study (2 timepoints with 6-month interval)	201 freshmen (mean age 14.16±0.44 years) from a local high school, of which 192 completed follow-up	Co-rumination with closest friend (27-item CRQ)	Outcomes measured: depressive symptoms assessed by Children's Depression Inventory Statistics: path analysis, allowing modeling of indirect effects (including rumination effects)
Van Zalk, 2017, Sweden	Observational: prospective cohort study (3 timepoints with 8-month intervals)	526 adolescents (423 initial participants and 103 of their online friends), aged 13-15 years (mean age 14.05, SD unknown), attending school in a medium-sized town in Sweden	Co-rumination with best online friend (8 questions from the revised CRQ)	Outcomes measured: - depressive symptoms: shortened version of the Child Depression Scale from the Center for Epidemiological Studies (CESD-10) - social anxiety: Social Phobia Screening Questionnaire for Children Statistics: autoregressive cross-lagged model
Webb, 2017, Australia	Observational: prospective cohort study (2 timepoints with 1-year interval)	365 students, aged 9-14 years (mean age 12.0±0.90 years), in grades 5, 6, or 7 attending one of three independent private schools in an urban area of Australia	- Appearance conversation with friends (6-item Appearance Conversations with Friends Scale) - Appearance teasing (Weight teasing subscale of the Perceptions of Teasing Scale)	Outcomes measured: appearance rejection sensitivity, assessed with Adolescent Appearance Rejection Sensitivity Scale

				Statistics: path model analysis

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
5-11 years				
Self-harm T4	Teased frequently (during one earlier wave vs no wave)	<u>Statistically significant</u>: aOR: 6.26, 95%CI [1.91;20.55] (p<0.05) *With harm for being teased frequently*	1, 273 vs 627	Borschmann, 2020
	Teased frequently (during 2 or 3 earlier timepoints vs no wave)	<u>Statistically significant</u>: aOR: 15.14, 95%CI [4.95;46.28] (p<0.05) *With harm for being teased frequently*	1, 116 vs 627	
12-18 years				
Co-dampening / co-enhancing				
Depressive symptoms (T1 --> T2)	Co-dampening (controlled for gender and co-rumination)	*General model* Not statistically significant: β: -0.05 £† (p>0.05) *Model including only adolescents with a stable friend* <u>Statistically significant</u>: β: -0.17 £ (p=0.001) *With benefit for co-dampening*	1, 543	Bastin, 2018a
	Co-dampening (controlled for gender, dampening and enhancing)	*General model* Not statistically significant: β: -0.06 £† (p>0.05) *Model including only adolescents with a stable friend* <u>Statistically significant</u>: β: -0.20 £ (p<0.001) *With benefit for co-dampening*		
Anhedonic symptoms (T1 --> T2)	Co-dampening (controlled for gender and co-rumination)	Not statistically significant: β: -0.03 £† (p>0.05)		
	Co-dampening (controlled for gender, dampening and enhancing)	Not statistically significant: β: -0.09 £† (p<0.10)		
Depressive symptoms (T1 --> T2)		Not statistically significant: β: 0.02 £† (p>0.05)		
	Co-enhancing (controlled for gender and co-rumination)	Not statistically significant: β: -0.01 £† (p>0.05)		
Anhedonic symptoms (T1 --> T2)	Co-enhancing (controlled for gender, dampening and enhancing)	Not statistically significant: β: -0.04 £† (p>0.05)		

	Co-enhancing (controlled for gender and co-rumination)	Not statistically significant: β: -0.03 £† (p>0.05)		

<table>
<tr><td colspan="5" style="background-color:#c5d9a0">Co-rumination</td></tr>
</table>

Depressive symptoms (T1 --> T2)	Co-rumination	*Twin difference analysis* Not statistically significant: β: 0.11, 95%CI [-0.07;0.28] £† (p>0.05) *Analysis of randomly selected member of each twin-pair* Not statistically significant: β: 0.05±0.08, 95%CI [-0.14; 0.21] £† (p=0.56)	1, 182 twin pairs §	Dirghangi, 2015
Depression (T --> T+1) (hierarchical linear model)		Statistically significant: B: 1.93±0.75 (p<0.01) *With harm for co-rumination*	1, 350 §	Hankin, 2010
Depression (T1 --> T2) (structural equation model)		Statistically significant: β: 0.15 £ (p<0.01) *With harm for co-rumination*		
Depressive symptoms (T1 --> T2)		Not statistically significant: β: 0.03 £† (p>0.05)	1, 192 §	Stone, 2015
		Not statistically significant: pr: 0.08 £† (partial correlation controlled for T1 depressive symptoms) (p>0.05)	1, 83 §	Starr, 2009
		Not statistically significant: β: -0.02 £† (p>0.05)	1, 526	Van Zalk, 2017
Depressive symptoms (T2 --> T3)		Statistically significant: β: 0.08 £ (p<0.05) *With harm for co-rumination*	1, 1205	Ohannessian, 2021
		Not statistically significant: β: 0.02 £† (p>0.05)	1, 526	Van Zalk, 2017
Depression contagion: friend depression (T1 --> T2)		Statistically significant: β: 0.25 £ (p<0.0001) *With harm for co-rumination*	1, 333 §	Schwartz-Mette, 2016
Depression contagion: adolescent depression (T1 --> T2)		Statistically significant: β: 0.25 £ (p<0.0001) *With harm for co-rumination*		
Onset of depressive episodes		Statistically significant: £ OR: 1.07 (p=0.003) *With harm for co-rumination*	1, 102 §	Stone, 2011
Depression onset		Statistically significant: £ OR: 1.07 (p=0.01) *With harm for co-rumination*	1, 95 §	Stone, 2011
Anhedonic depression (T --> T+1)		Statistically significant: B: 0.045±0.02 (p<0.05) *With harm for co-rumination*	1, 350 §	Hankin, 2010

Anxiety (T1 --> T2)		*Twin difference analysis* <u>Statistically significant</u>: β: 0.19, 95%CI [0.04;0.35] (p<0.05) *With harm for co-rumination* *Analysis of randomly selected member of each twin-pair* <u>Statistically significant</u>: β: 0.14±0.07, 95%CI [0.01;0.28] (p=0.05) *With harm for co-rumination*	1, 182 twin pairs §	Dirghangi, 2015
Anxiety symptoms (T1 --> T3)	Co-rumination (T2)	<u>Statistically significant</u>: β: 0.14 £ (p<0.001) *With harm for co-rumination*	1, 1205	Ohannessian, 2021
Anxious arousal (T --> T+1)	Co-rumination	Not statistically significant: B: 0.12±0.06 £† (p<0.06)	1, 350 §	Hankin, 2010
Social anxiety (T1 --> T2)		Not statistically significant: β: 0.03 £† (p>0.05)	1, 526	Van Zalk, 2017
Social anxiety (T2 --> T3)		Not statistically significant: β: 0.05 £† (p>0.05)		
General internalizing symptoms (T --> T+1)		<u>Statistically significant</u>: B: 0.09±0.02 (p<0.001) *With harm for co-rumination*	1, 350 §	Hankin, 2010
General externalizing symptoms (T --> T+1)		Not statistically significant: B: 0.009±0.03 £† (p>0.05)		
Self-harm (T1 --> T2)		<u>Statistically significant</u>: β: -0.07 £ (p<0.05) *With benefit for co-rumination*	1, 709	Latina, 2015
Approval of substance use				
Marijuana use (T1 --> T2)	Approval of marijuana use by peers	Not statistically significant: £ OR: 1.1, 95%CI [0.9;1.4] ¥ (p=0.25)	1, 1713	Ellickson, 2004
Marijuana use (T2 --> T3)		Not statistically significant: £ OR: 1.1, 95%CI [0.8;1.5] ¥ (p=0.46)	1, 1590	
Marijuana use (T3 --> T4)		<u>Statistically significant</u>: £ OR: 1.5, 95%CI [1.0;2.1] (p=0.03) *With harm for approval of marijuana use by peers*	1, 1557	
Friend communication				
Depression (T1 --> T2)	Secrecy from friends	Not statistically significant: B: -0.13±0.15 £† (p=0.40)	1, 148 §	Laird, 2013
Antisocial behaviour (T1 --> T2)		Not statistically significant: B: -0.05±0.03 £† (p=0.14)		
Changes in weight concern	Peer teasing	*Girls:* Not statistically significant: B: 0.07±0.12 £† (p>0.05) *Boys:*	1, 236 § 1, 193 §	Helfert, 2011

		Not statistically significant: B: 0.11±0.09 £† (p>0.05)		
Body dissatisfaction (T1 --> T2)	Appearance conversations	*Girls:* Not statistically significant: β: 0.06 £† (p>0.05) *Boys:* Not statistically significant: β: 0.05 £† (p>0.05)	1, 164 § 1, 139 §	Jones, 2004
Body dissatisfaction (T1 --> T2)	Appearance teasing	*Girls:* Not statistically significant: £† (ns in univariate analysis and therefore not included in multivariate analysis) (p>0.05) *Boys:* Not statistically significant: β: 0.09 £† (p>0.05)	1, 164 § 1, 139 §	Jones, 2004
Body dissatisfaction (T1 --> T2)	Weight teasing	*Girls:* Not statistically significant: β: -0.54±0.46 £† (p>0.05) *Boys:* Statistically significant: β: 1.48±0.50 (p≤0.01) *With harm for weight teasing*	1, 440 1, 366 §	Paxton, 2006
5-18 years				
Co-rumination				
Depressive symptoms (T1 --> T2)	Co-brooding	Statistically significant: B: 0.20±0.09 (p<0.05) *With harm for co-brooding*	1, 357 §	Bastin, 2014
		Boys and girls Not statistically significant: β: 0.05 £† (p>0.05) *Girls only* Not statistically significant: β: 0.16 £† (p=0.06) *Boys only* Not statistically significant: β: -0.18 £† (p=0.18)	1, 313 (boys and girls) §	Bastin, 2018b
		Not statistically significant: £† (p>0.05)	1, 1188	Bastin, 2021
Depressive symptoms (T2 --> T3)		Not statistically significant: £† (p>0.05)	1, 899	
Brooding rumination (T1-->T2)		Statistically significant: £† (p<0.05) *With harm for co-brooding*	1, 1188	

Brooding rumination (T2-->T3)		<u>Statistically significant:</u> £† (p<0.05) *With harm for co-brooding*	1, 899	
Depressive symptoms (T1 --> T2)	Co-reflection	<u>Statistically significant:</u> B: -0.30±0.10 (p<0.01) *With benefit for co-reflection*	1, 357 §	Bastin, 2014
		Boys and girls Not statistically significant: β: -0.06 £† (p>0.05) *Girls only* <u>Statistically significant:</u> β: -0.17 £ (p=0.04) *With benefit for co-reflection* *Boys only* Not statistically significant: β: 0.17 £† (p=0.20)	1, 313 (boys and girls) §	Bastin, 2018b
Depressive symptoms (in function of interpersonal stress (T1 --> T4)	Co-rumination	Not statistically significant: B: -0.05±0.05 £† (p>0.05)	1, 291 §	Bastin, 2015
Depressive symptoms (in function of non-interpersonal stress) (T1 --> T4)		Not statistically significant: B: -0.01±0.14 £† (p>0.05)	1, 291 §	
Depressive symptoms (T1 --> T3)		Not statistically significant: β: -0.04 £† (p=0.13)	1, 899	Bastin, 2021
Depression (T1 --> T2)		*Boys and girls* <u>Statistically significant:</u> β: 0.07 £ (p<0.05) *With harm for co-rumination* *Girls* <u>Statistically significant:</u> β: 0.11 £ (p<0.01) *With harm for co-rumination* *Boys* Not statistically significant: β: 0.01 £† (p>0.05)	1, 604 1, 322 § 1, 282 §	Rose, 2007
Brooding rumination (T1 --> T2)		<u>Statistically significant:</u> β: 0.05 £ (p<0.001) *With harm for co-rumination*	1, 1188	Bastin, 2021
Brooding rumination (T2 --> T3)		<u>Statistically significant:</u> β: 0.06 £ (p<0.001) *With harm for co-rumination*	1, 899	
Anxiety (T1 --> T2)		*Boys and girls* Not statistically significant: β: 0.03 £† (p>0.05)	1, 604	Rose, 2007

		Girls Statistically significant: β: 0.10 £ (p<0.05) *With harm for co-rumination*	1, 322 §	
		Boys Not statistically significant: β: -0.04 £† (p>0.05)	1, 282 §	
Peer communication				
Marijuana use (T1 --> T4)	Targeted best-friend communication against drugs at T2	Not statistically significant: £† (p>0.05)	1, 2749	Kam, 2013
Inhalant use (T1 --> T4)		Not statistically significant: £† (p>0.05)		
Ecstasy use (T1 --> T4)		Statistically significant: β: 0.19 £ (p<0.05) *With harm for targeted best friend communication*		
Marijuana use (T1 --> T3)	Targeted best-friend communication against marijuana use	Not statistically significant: β: 0.04 £† (p>0.05)	1, 627	Kam, 2015
Bullying perpetration (T1 --> T2)	Sexting: sending	Not statistically significant: B: -0.08±0.14 £† (p=0.589)	1, 1456	Ojeda, 2019
Cyberbullying (T1 --> T2)		Statistically significant: B: 0.26±0.13 (p=0.048) *With harm for sending sexts*		
Bullying perpetration (T1 --> T2)	Sexting: receiving	Statistically significant: B: 0.24±0.10 (p=0.016) *With harm for receiving sexts*		
Cyberbullying (T1 --> T2)		Not statistically significant: B: 0.16±0.12 £† (p=0.129)		
Appearance rejection sensitivity (T1 --> T2)	Peer appearance conversations	Not statistically significant: B: 0.73±0.42, 95%CI [0.31;1.15] £† (p=0.08)	1, 365 §	Webb, 2017
Appearance rejection sensitivity (T1 --> T2)	Appearance teasing	Not statistically significant: B: -0.01±0.03, 95%CI [-0.08;0.05] £† (p=0.70)	1, 365 §	Webb, 2017
Body dissatisfaction (T1 --> T2)	Peer discussion about dieting	Statistically significant: path coefficient B: 0.46 £ (p<0.05) *With harm for peer discussion about dieting*	1, 85 §	Blodgett Salafia, 2010

B ± SE (unless otherwise indicated), β ± SE (unless otherwise indicated), OR: odds ratio, SE: standard error, B: unstandardized coefficient, β: standardized coefficient, pr: partial correlation coefficient
£ No SE and/or CI available; or no information on magnitude of effect available to assess variability of results
¥ Imprecision (large variability of results)
† Imprecision (lack of data)
§ Imprecision (limited sample size or low number of events)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Bastin, 2014	Yes, community sample of 9 schools, no information on how these schools were selected.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline depressive symptoms, gender, intrapersonal brooding and reflection.	No, 96.2% of the initial sample was available at follow-up; also expectation maximization (EM) algorithm was used to impute scale-based missing data at T2 which enabled analysis of the full T1 sample.	No, no indication
Bastin, 2015	Yes, community sample of 9 schools, no information on how these schools were selected.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline depressive symptoms, stress, gender and brooding	Yes, 77 of 368 pupils did not complete all assessments; there was a significant difference in baseline depression symptoms between those who did and did not complete all assessments.	No, no indication
Bastin, 2018a	Yes, community sample of 7 schools, no information on how these schools were selected.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline symptoms in 2 models, either gender and co-rumination, or gender, dampening and enhancing.	Yes, 18% attrition; drop-outs had higher baseline scores on several variables.	Yes, subgroup analysis of stable friendships were done post hoc and not available for all risk factors and outcomes.
Bastin, 2018b	Yes, community sample of 11 schools, no information on how these schools were selected. Also, only participants with best friend at both timepoints were included in the analysis.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline symptoms and age.	Yes, 97.8% of the T1 sample was available at T2, but of those, only participants naming the same person as their best friend at both assessments were included in the analyses and this group differed from the original T1 sample in baseline depressive symptoms.	No, no indication
Bastin, 2021	No, schools were randomly	Yes, exposure and outcome	No, multivariable	Yes, participants with higher	No, no indication

	selected and all students were invited to participate.	measures were self-reported.	analysis accounted for baseline co-brooding rumination, depressive symptoms, age and gender.	levels of depressive symptoms, participants with higher brooding rumination, and boys, were more likely to drop out of the study.	
Blodgett Salafia, 2010	Yes, participants recruited in primary schools in one medium-sized, Midwestern city. Sample consisted of primarily European American early adolescent girls who came from upper-middle-class families.	Yes, exposure and outcome measures were self-reported.	No, path analysis accounted for parental discussion of dieting and bulimic symptoms at baseline.	No, no significant differences between the 85 mother-girl dyads who participated and those who did not.	No, no indication
Borschmann, 2020	No, participants came from a stratified random sample of 43 primary schools.	Yes, outcome and exposure measures were self-reported.	No, adjusted for age, sex and SEIFA (Socio-Economic Index For Areas) quintile.	Unclear, no information on differences between those who did all 4 timepoints, and those who did not.	Yes, the sample was skewed towards higher SES and had a higher proportion of Indigenous children than the general Australian population. They did not ask about self-harm prior to T4 and this may represent a missed opportunity to collect informative data.
Dirghangi, 2015	Yes, only monozygotic twins were included in the study.	Yes, exposure and outcome measures were self-reported; the short version of the co-rumination scale may have tapped aspects of the construct that are more closely aligned with anxiety than with depression.	No, multivariable analysis accounted for baseline symptoms.	No, of the 266 MZ twin pairs originally recruited at birth, 133 (50%) completed questionnaires at T1 and T2, 24 (9.0%) completed questionnaires at T1 only, and 25 (9.4%) completed questionnaires at T2 only. There were no significant differences between the	No, no indication

				groups. Imputed data sets were generated.	
Ellickson, 2004	No, part of a larger study on drug prevention. Schools were drawn from 8 school districts, to represent highly diverse environments.	Yes, exposure and outcome measures were self-reported, although outcomes were confirmed by saliva sample.	No, multivariable analysis accounted for demographics and several other variables from the year before.	Yes, 20% loss of follow up, which necessitated the use of sample weights to effectively reduce attrition-related bias, and the use of somewhat different subsamples of participants across analyses.	No, no indication
Hankin, 2010	Yes, participants recruited from 5 Chicago area schools, no random sampling.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline symptoms.	No, there was loss of follow-up but attrition analyses showed that missing data were not related to demographic characteristics (age, sex, ethnicity, or any initial symptom or stressor scores). Data were thus viewed as missing at random for analyses.	No, no indication
Helfert, 2011	Yes, the study was conducted on middle-and upper-class students in grades 7–9 in 3 German high schools that cooperate with the author's institution for different research projects.	Yes, exposure and outcome measures were self-reported.	No, hierarchical regression analysis accounted for BMI, age, parental teasing, and other variables.	Yes, participants who dropped out were significantly older and had a higher BMI.	Schools received financial incentives for their participation.
Jones, 2004	Yes, no random sampling, participants came from only 4 schools in one district.	Yes, exposure and outcome were self-reported.	No, multivariate multiple regression analysis.	No, no differences between participants and drop-outs.	No, no indication
Kam, 2013	No, part of a larger prospective study; one child per household was randomly selected.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for prior drug use, sex, ethnicity/ race, age, yearly household income, and	Unclear, drop-out not described.	No, no indication

			parent's education level.		
Kam, 2015	Yes, students recruited from 3 rural Illinois public schools, no random sampling.	Yes, exposure and outcome measures were self-reported. Also communication was not assessed for each substance, although conversation topics may differ by substance.	No, multivariable analysis accounted for time spent in the United States, age, school grades, religiosity, and past marijuana use.	Yes, students who left the study at T2 had significantly weaker anti-substance use beliefs and greater marijuana use than students who were at all three timepoints.	No, no indication
Laird, 2013	Yes, participating families were recruited from 1 public school, no random sampling.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for earlier adjustment, secrecy from parents, and the quality of the friendship.	Yes, 18.3% loss of follow-up; girls were more likely to drop out; dropouts had higher quality friendships at age 12.	Participants were compensated 70$ per family at T1 and 90$ per family at T2.
Latina, 2015	No, schools were randomly selected and all students were invited to participate.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline self-harm, depressive symptoms, communication and age.	No, there was some attrition (15%), but there were no differences between the youth who participated at both time points and those who did not participate at the second data collection.	Yes, assessments of self-harming behaviors were made at a six-month interval while the self-harming behaviors referred to were over a one-year period, thereby creating an overlap.
Ohannessian, 2021	Yes, students from only 5 schools invited, no random sampling.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline depressive symptoms, anxiety and social media use	Unclear, not mentioned whether drop-out and no attrition analyses	No, no indication
Ojeda, 2019	Yes, opportunity sampling of schools and young people.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline bullying and sexting behaviour, gender and age.	Unclear, 16.1% loss of follow-up, but unclear whether some students were more likely to drop out than others.	Yes, bullying episodes during the period between data collection points may have been missed.
Paxton, 2006	Unclear, participants recruited from 31 Minnesota	Yes, exposure and outcome were self-reported.	No, multivariate linear regression	Yes, there were small but significant differences	No, no indication

	schools, unclear if sampling was random.		analyses performed taking into BMI and demographic variables into account.	between participants at T1 and dropouts at T2.	
Rose, 2007	Yes, participants recruited from 4 schools, no random sampling.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline symptoms.	Yes, 5% loss of follow-up and attrition analysis showed that youths who participated in both assessments scored lower on depression than youths who only participated at T1.	No, no indication
Schwartz-Mette, 2016	Yes, data from larger project. Participants selected based on availability of data. No random sampling.	Yes, exposure and outcome measures were self-reported; T1 data were collected in the lab and T2 data in the lab or via mail.	No, multivariable analysis accounted for baseline symptoms	Yes, 31% loss of follow-up and difference in friendship quality between those who participated at both time points and those who participated only at T1.	No, no indication
Starr, 2009	Yes, girls were recruited either from a previous study or by flyer, and there was a difference in baseline depressive symptoms between the two groups.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline depressive symptoms.	Yes, 12% of girls did not participate at T2. Girls who did not participate at T2 showed more depressive symptoms ($p<0.05$), and thus longitudinal results may underestimate actual effects.	Yes, bias in outcome reporting: the regression coefficients of the hierarchical model were only available for the interaction term, not for the main effects.
Stone, 2011	Yes, no random sampling, participants recruited through ads in local newspaper.	Yes, exposure was self-reported, although outcome was assessed by a trained clinician.	No, multivariable analysis accounted for baseline depressive symptoms and rumination.	No, complete data were available over the full 2 years for all but one family.	No, no indication
Stone, 2015	Yes, students from one local high school, no random sampling, opportunity sampling of school and students.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline symptoms.	No, missing data at each assessment is limited (6 to 7%). It was examined if data were missing at random to justify data imputation methods for estimating missing values. Little's missing completely at random (MCAR)	No, no indication

				test was non-significant, supporting the imputation of missing values.	
Van Zalk, 2017	Yes, no random sampling, participants from one medium-sized town in Sweden.	Yes, exposure and outcome measures were self-reported; also at T1-T2 information was collected through a combination of offline and online questionnaires, whereas at T3 all questionnaires were online-only. For online friends all questionnaires were taken online.	No, multivariable analysis accounted for baseline social anxiety, depressive symptoms, friendship quality and friendship stability.	Unclear, 72% of the adolescents reported data on all study variables at all three time points, but no attrition analysis.	No, no indication
Webb, 2017	Yes, no random sampling, participants recruited in 3 private schools.	Yes, exposure and outcome measures were self-reported.	No, multivariable analysis accounted for baseline appearance rejection sensitivity, BMI percentile, general rejection sensitivity, and peer-reported attractiveness.	No, 6% loss of follow-up but no differences in demographics or T1 measures between those who did, and those who did not complete the T2 assessment.	Yes, measuring participants' height and weight during the session in which they completed the questionnaire may have primed appearance-related concerns.

Certainty of the body of evidence

5-11 years

	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

12-18 years

Co-dampening / co-enhancing	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	

		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Co-rumination	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Approval of substance use	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	Large variability of results only in non-significant outcomes
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Friend communication	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

5-18 years

Co-reflection	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Co-brooding/co-rumination	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	Limited sample sizes/lack of data only for non-significant results
Inconsistency	0	
Indirectness	0	
Publication bias	0	

		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Targeted best friend communication against drugs	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	Lack of data only for non-significant results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Sexting	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	Lack of data only for non-significant results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Peer discussion about dieting/appearance rejection sensitivity	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample size
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	*5-11 years* There is limited evidence with harm for **being teased frequently**. It was shown that being teased frequently in the past resulted in a statistically significant increased risk of self-harm, compared to not being teased frequently in the past (Borschmann 2020). Evidence is of very low certainty. *12-18 years* *Co-dampening, co-enhancing and co-rumination* There is limited evidence with benefit for **co-dampening** (talking about positive emotions in a condescending manner within a dyadic relationship). It was shown that co-dampening with a stable friend resulted in a statistically significant decreased risk of depressive symptoms (Bastin 2018a). However, in a general model (all types of friends), a statistically significant decreased risk of depressive or anhedonic symptoms in case of co-dampening could not be demonstrated (Bastin 2018a).

Evidence is of very low certainty and results cannot be considered precise due to lack of data.

There is limited evidence concerning the risk of depressive or anhedonic symptoms in presence of **co-enhancing** (elaborating the positive aspects of positive emotions within a dyadic relationship).
A statistically significant decreased risk of depressive or anhedonic symptoms in case of co-enhancing could not be demonstrated (Bastin 2018a).
Evidence is of very low certainty and results cannot be considered precise due to lack of data.

There is limited evidence with harm for **co-rumination** (discussing personal problems within a dyadic relationship, in the form of mutual encouragement of problem talk or dwelling on negative affects).
It was shown that co-rumination resulted in a statistically significant increased risk of (anhedonic) depression, depression onset, depressive symptoms or depression contagion (Hankin 2010, Ohannessian 2021, Schwartz-Mette 2016, Stone 2011). However, in 4 other studies, a statistically significant decreased risk of depression or depressive symptoms in case of co-rumination could not be demonstrated (Dirghangi 2015, Starr 2009, Stone 2015, Van Zalk 2017).
It was shown that co-rumination resulted in a statistically significant increased risk of anxiety or anxiety symptoms (Dirghangi 2015, Ohannessian 2021). However, a statistically significant decreased risk of anxious arousal, social anxiety in case of co-rumination could not be demonstrated in 2 other studies (Hankin 2010, Van Zalk 2017).

It was shown that co-rumination resulted in a statistically significant increased risk of general internalizing symptoms. On the other hand, in the same study, a statistically significant decreased risk of general externalizing symptoms in case of co-rumination could not be demonstrated (Hankin 2010).

In one study, it was shown that co-rumination resulted in a statistically significant decreased risk of self-harm (Latina 2015).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data.

Approval of substance use
There is limited evidence with harm for **approval of marijuana use by peers**.
It was shown that approval of marijuana use by peers resulted in a statistically significant increased risk of marijuana use in students of 10th grade (Ellickson 2004).
In younger students (8th – 9th grade), an increased risk of marijuana use in case of approval of marijuana use by peers could not be demonstrated (Ellickson 2004).
Evidence is of very low certainty.

Friend communication
There is limited evidence with harm for **weight teasing** in boys.
It was shown that weight teasing resulted in a statistically significant increased risk of body dissatisfaction in boys. However, in girls, an increased risk of body satisfaction in case of weight teasing could not be demonstrated (Paxton 2006).
Evidence is of very low certainty and results could not be considered precise due to limited sample size and/or lack of data.

There is limited evidence concerning the risk of depression, antisocial behaviour, changes in weight concern or body dissatisfaction in case of **secrecy from friends**, **peer teasing**, **appearance conversations** or **appearance teasing**.
A statistically significant increased risk of depression or antisocial behaviour in case of secrecy from friends could not be demonstrated (Laird 2013). In addition, an increased risk of changes in weight concern in case of peer teasing could not be demonstrated (Helfert 2011). Also, a statistically significant increased risk of body dissatisfaction in case of appearance conversations or appearance teasing could not be demonstrated (Jones 2004).
Evidence is of very low certainty and results could not be considered precise due to limited sample size and lack of data.

5-18 years

	Co-rumination, co-brooding, co-reflection There is limited evidence with benefit for **co-reflection** (the tendency to try to better understand what is happening and make causal analyses within a dyadic relationship; specific items of the Co-Rumination Questionnaire). It was shown that co-reflection resulted in a statistically significant decrease in depressive symptoms (Bastin 2014, 2018b (girls only)). However, a statistically significant decrease in depressive symptoms, in case of co-reflection, could not be demonstrated in boys and girls together, or in girls alone (Bastin 2018b). Evidence is of very low certainty and results could not be considered precise due to limited sample size and lack of data. There is limited evidence with harm for **co-brooding** (the tendency to catastrophize and focus attention on negative, undesirable feelings and consequences of problems within a dyadic relationship; specific items from the Co-Rumination Questionnaire) and **co-rumination**. It was shown that co-brooding resulted in a statistically significant increase in depressive symptoms (Bastin 2014) and brooding rumination (Bastin 2021). However, a statistically significant increase in depressive symptoms, in case of co-brooding, could not be demonstrated in two studies (Bastin 2018b, 2021). It was shown that co-rumination resulted in a statistically significant increase in depression (Rose 2007). However, a statistically significant increased risk in depressive symptoms could not be demonstrated in 2 other studies (Bastin 2021, Bastin 2015), and a statistically significant increased risk in depression in boys alone could also not be demonstrated (Rose 2007). In addition it was shown that co-rumination resulted in a statistically significant increase in brooding rumination (Bastin 2021) and anxiety in girls (Rose 2007). A statistically significant increased risk anxiety in boys and girls, or boys alone (Rose 2007) in case of co-rumination could not be demonstrated . Evidence is of very low certainty. *Peer communication* There is limited evidence with harm for **targeted best friend communication against drugs**. It was shown that targeted best friend communication against drugs resulted in a statistically significant increased risk of ecstasy use (Kam 2013). On the other hand, a statistically significant increased risk of marijuana use or inhalant use, in case of targeted best friend communication against drugs, could not be demonstrated (Kam 2013, 2015). Evidence is of very low certainty. There is limited evidence with harm for **sending** or **receiving sexts.** It was shown that sending sexts resulted in a statistically significant increased risk of cyberbullying. In addition, it was shown that receiving sexts resulted in a statistically significant increased risk of bullying perpetration. However, a statistically significant increased risk of bullying perpetration in case of sending sexts, or an increased risk of cyberbullying in case of receiving sexts could not be demonstrated (Ojeda 2019). Evidence is of very low certainty. There is limited evidence with harm for **peer discussion about dieting**. It was shown that peer discussion about dieting resulted in a statistically significant increased risk of body dissatisfaction (Blodgett Salafia 2010). Evidence is of very low certainty and results cannot be considered precise due to limited sample size. There is limited evidence concerning the risk of appearance rejection sensitivity in case of **peer appearance conversation or appearance teasing**. A statistically significant increased risk of appearance rejection sensitivity in case of peer appearance conversation or appearance teasing could not be demonstrated (Webb 2017). Evidence is of very low certainty and results cannot be considered precise due to limited sample size.
Reference(s)	**Articles** Bastin M, Bijttebier P, Raes F, Vasey MW. *Brooding and reflecting in an interpersonal context*. Personality and individual differences 2014, 63:100-105.

Bastin M, Mezulis AH, Ahles J, Raes F, Bijttebier P. *Moderating effects of brooding and co-rumination on the relationship between stress and depressive symptoms in early adolescence: A multi-wave study*. Journal of Abnormal Child Psychology 2015, 43(4):607-618.

Bastin M, Nelis S, Raes F, Vasey MW, Bijttebier P. *Party pooper or life of the Party: dampening and enhancing of positive affect in a peer context*. Journal of abnormal child psychology 2018a, 46(2):399-414.

Bastin M, Vanhalst J, Raes F, Bijttebier P. *Co-brooding and co-reflection as differential predictors of depressive symptoms and friendship quality in adolescents: Investigating the moderating role of gender*. Journal of youth and adolescence 2018b, 47(5):1037-1051.

Bastin M, Luyckx K, Raes F, Bijttebier P. *Co-Rumination and Depressive Symptoms in Adolescence: Prospective Associations and the Mediating Role of Brooding Rumination*. Journal of youth and adolescence 2021, 50(5):1003-1016.

Blodgett Salafia EH and Gondoli DM. *A 4-Year Longitudinal Investigation of the Processes by Which Parents and Peers Influence the Development of Early Adolescent Girls' Bulimic Symptoms*. Journal of Early Adolescence 2011, 31(3):390–414

Borschmann R, Mundy LK, Canterford L, Moreno-Betancur M, Moran PA, Allen NB, Viner RM, Degenhardt L, Kosola S, Fedyszyn I, Patton GC. *Self-harm in primary school-aged children: Prospective cohort study.* PLoS One 2020, 15(11):e0242802.

Dirghangi S, Kahn G, Laursen B, Brendgen M, Vitaro F, Dionne G, Boivin M. *Co-rumination cultivates anxiety: A genetically informed study of friend influence during early adolescence*. Developmental psychology 2015, 51(4):564.

Ellickson PL, Tucker JS, Klein DJ, Saner H. *Antecedents and outcomes of marijuana use initiation during adolescence*. Preventive medicine 2004, 39(5):976-984.

Hankin BL, Stone L, Wright PA. *Corumination, interpersonal stress generation, and internalizing symptoms: Accumulating effects and transactional influences in a multiwave study of adolescents*. Development and psychopathology 2010, 22(1):217-235.

Helfert S and Warschburger P. *A prospective study on the impact of peer and parental pressure on body dissatisfaction in adolescent girls and boys*. Body Image 2011, 8:101-109

Jones DC. *Body Image Among Adolescent Girls and Boys: A Longitudinal Study*. Developmental Psychology 2004, 40(5):823–835

Kam JA, Lee C-j. *Examining the effects of mass media campaign exposure and interpersonal discussions on youth's drug use: The mediating role of visiting pro-drug websites*. Health Communication 2013, 28(5):473-485.

Kam JA, Wang N. *Longitudinal effects of best-friend communication against substance use for Latino and non-Latino White early adolescents*. Journal of research on adolescence 2015, 25(3):534-550.

Laird RD, Bridges BJ, Marsee MA. *Secrets from friends and parents: longitudinal links with depression and antisocial behavior*. J Adolesc 2013, 36(4):685-693.

Latina D, Giannotta F, Rabaglietti E. *Do friends' co-rumination and communication with parents prevent depressed adolescents from self-harm?* Journal of Applied Developmental Psychology 2015, 41:120-128.

Ohannessian CM, Fagle T, Salafia C. *Social media use and internalizing symptoms during early adolescence: The role of co-rumination*. J Affect Disord 2021, 280(Pt A):85-88.

Ojeda M, Del Rey R, Hunter SC. *Longitudinal relationships between sexting and involvement in both bullying and cyberbullying*. Journal of adolescence 2019, 77:81-89.

Paxton SJ, Eisenberg ME, Neumark-Sztainer D. *Prospective Predictors of Body Dissatisfaction in Adolescent Girls and Boys: A Five-Year Longitudinal Study*. Developmental Psychology 2006, 42(5):888-899

Rose AJ, Carlson W, Waller EM. *Prospective associations of co-rumination with friendship and emotional adjustment: considering the socioemotional trade-offs of co-rumination*. Developmental psychology 2007, 43(4):1019.

Schwartz-Mette RA, Smith RL. *When does co-rumination facilitate depression contagion in adolescent friendships? Investigating intrapersonal and interpersonal factors*. Journal of Clinical Child & Adolescent Psychology 2018, 47(6):912-924.

Starr LR, Davila J. *Clarifying co-rumination: associations with internalizing symptoms and romantic involvement among adolescent girls*. J Adolesc 2009, 32(1):19-37.

Stone LB, Hankin BL, Gibb BE, Abela JRZ. *Co-Rumination Predicts the Onset of Depressive Disorders During Adolescence*. Journal of Abnormal Psychology 2011, 120(3):752-757.

Stone LB, Gibb BE. *Brief report: preliminary evidence that co-rumination fosters adolescents' depression risk by increasing rumination.* J Adolesc 2015, 38:1-4.
Van Zalk N, Tillfors M. *Co-rumination buffers the link between social anxiety and depressive symptoms in early adolescence.* Child and Adolescent Psychiatry and Mental Health 2017, 11(1).
Webb HJ, Zimmer-Gembeck MJ, Waters AM, Farrell LJ, Nesdale D, Downey G. *"Pretty pressure" from peers, parents, and the media: A longitudinal study of appearance-based rejection sensitivity.* Journal of Research on Adolescence 2017, 27(4):718-735.

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Co-Rumination Questionnaire (CRQ)			
Co-rumination: discuss personal problems within a dyadic relationship, in the form of mutual encouragement of problem talk or dwelling on negative affects	Bastin, 2015 Rose, 2007 Schwartz-Mette, 2016 Starr, 2009 Stone, 2015	Co-Rumination Questionnaire (CRQ)	A 27-item self-report inventory, measuring the extent to which participants co-ruminated with friends. The measure covers nine content areas, with three items each: (1) frequency of problem discussion ("We spend most of our time together talking about problems that my friend or I have"), (2) tendency to talk about problems rather than doing other activities ("If one of us has a problem, we will talk about the problem rather than talking about something else or doing something else"), (3) how much the adolescent encourages her friends to discuss problems ("After my friend tells me about a problem, I always try to get my friend to talk more about it later"), (4) how much the adolescent's friends encourage her to discuss problems ("When I have a problem, my friend always tries really hard to keep me talking about it"), (5) tendency to repeatedly revisit the same problems ("When we talk about a problem that one of us has, we will talk about every part of the problem over and over"), (6) debate about potential causes of problems ("We talk for a long time trying to figure out all the different reasons why the problem might have happened"), (7) estimating the consequences of problems ("We try to figure out every one of the bad things that might happen because of the problem"), (8) conjecture about aspects of the problem that are not understood ("We spend a lot of time trying to figure out parts of the problem that we cannot understand"), and (9) heightened focus on negative emotions ("We talk a lot about how bad the person with the problem feels"). Each item was rated on a Likert-type scale, ranging from 1 ("not at all true") to 5 ("really true"). Mean score of all 27 items can be used to compute an overall co-rumination score.
Co-brooding: the tendency to catastrophize and focus attention on negative, undesirable feelings and consequences of problems within a dyadic relationship	Bastin, 2014	Selected items from the Co-Rumination Questionnaire (CRQ)	6 'co-brooding' items (e.g., "When we talk about a problem that one of us has, we try to figure out every one of the bad things that might happen because of the problem") and
Co-reflection: the tendency to try to better understand what is happening			5 'co-reflection' items (e.g., "When we talk about a problem that one of us has, we talk about all of the reasons why the problem might have

			happened"). Items are rated on a 5-point rating scale (1 = not at all true to 5 = really true).
and make causal analyses within a dyadic relationship			
Co-rumination (co-brooding and co-reflection)	Bastin, 2018b	12-item abbreviated version of the Co-Rumination Questionnaire (CRQ)	12 items of the CRQ were used, with six items assessing co-brooding (e.g., "When we talk about a problem that one of us has, we try to figure out every one of the bad things that might happen because of the problem") and six items assessing co-reflection (e.g., "When we talk about a problem that one of us has, we talk about all of the reasons why the problem might have happened"). Items are rated on a 5-point rating scale (1 = not at all true to 5 = really true).
Co-rumination	Bastin, 2021 Hankin, 2010 Stone, 2011 Latina, 2015	9-item abbreviated version of the Co-Rumination Questionnaire (CRQ)	One item for each of the 9 content areas (frequency of discussing problems, discussing problems instead of engaging in other activities, friend encouraging discussion of problems, target child encouraging friend to discuss problems, discussing the same problem repeatedly, speculation about causes, speculation about consequences, speculation about parts of the problem that are not understood, and focusing on negative feelings) of co-rumination were used. Co-rumination scores represent the means of all items.
Co-rumination	Dirghangi, 2015	4-item abbreviated version of the Co-Rumination Questionnaire (CRQ)	Rating the frequency of 4 items ("When my friend and I talk about a problem that one of us has, we'll talk about every part of the problem over and over"; "When my friend and I talk about a problem that one of us has, we talk for a long time trying to figure out all of the reasons why the problem might have happened"; "When my friend and I talk about a problem that one of us has, we talk a lot about all the different bad things that might happen because of the problem"; "When my friend and I talk about a problem that one of us has, we talk a lot about how bad the person with the problem feels") on a scale ranging from 1 (not at all true) to 5 (very true). The items selected represented the following conceptual factors: discussing the same problem repeatedly; speculation about causes; speculation about consequences; and focusing on negative feelings. Item scores were summed.
Co-rumination	Ohannessian, 2021	Co-Rumination Questionnaire, Short Version (CRQ-S)	Participants were asked to think about the way they usually are with their best or closest friends. A representative item is, "We talk about every part of the problem over and over." The response scale ranges from 1 = not at all true to 5 = really true. The five CRQ-S items can be averaged to reflect a rumination score. Unclear how many and which items of the CRQ were used.
Co-rumination with best online friend	Van Zalk, 2017	Revised co-rumination questionnaire	Eight questions about co-rumination from the revised co-rumination questionnaire. The original revised version used questions about adolescents' co-rumination about their problems with their mothers. In this study, they instead measured how the target participants talk about their problems with their best friends. The items measured to what extent the adolescents typically co-ruminated about when they have a problem and how they and their best friend usually talk about it. Examples of items were: When I have a problem, "my friend and I talk to each other about it for a long time," "we'll talk about every part of the problem over and

			over," and "we talk a lot about all of the different bad things that might happen because of the problem." The response items were (1) Not at all true, (2) A little true, (3) Somewhat true, (4) Mostly true, and (5) Really true.
Co-Dampening and Co-Enhancing Questionnaire (CoDEQ)			
Co-dampening: talking about positive emotions in a downgrading manner within a dyadic relationship Co-enhancing: elaborating on the positive aspects of positive emotions within a dyadic relationship	Bastin, 2018a	Co-Dampening and Co-Enhancing Questionnaire (CoDEQ)	Respondents have to indicate how often they respond in the described way when one of them feels glad or happy and they are talking about this. The rating scale has four response options: almost never (1), sometimes (2), often (3), and almost always (4). Eighteen items were developed; nine items intended to measure co-enhancing and nine items intended to measure co-dampening. Co-dampening: the following responses to positive affect are described: thinking about the fleetingness of positivity, thinking about worries, focusing on negative aspects of the positive affect or event, making upward social comparisons (i.e., how others are even better off than you), making external attributions (e.g., thinking "it was just luck"), and starting to think about past negative events. Co-enhancing: behavioral display, focusing on positive feelings, (e.g., thinking about how energetic one feels), thinking about positive past and future events, making downward social comparisons (i.e., comparing yourselves to those who are less fortunate), and thinking about positive self-qualities such as the ability to achieve whatever you desire.
Other scales or non-validated scales			
Best-friend communication against drugs use	Kam, 2013	Not a validated scale	Youth who reported that they had talked to their friends about drugs in the last 6 months were asked: "In the last 6 months, what sorts of things have you and your friends talked about? We talked about …" to which youth responded to the following segments: "specific things I could do to stay away from drugs" and "bad things that happen if you use drugs." Youth also were asked if they "talked with friends about anti-drug ads in recent months." These three variables were dichotomous (0=no, 1=yes). The responses of youth who reported "never" to how often they talked to parents or friends about drugs were re-coded as "no" for the targeted friend-to-friend communication.
Best-friend communication against substance use	Kam, 2015	6 items from the Targeted Child Communication about Alcohol Scale	Items were adapted to reflect on a best friend and to incorporate more substances. Students were asked, "How often has your best friend told you the following?" Sample questions were, "This person has…" "… told me they would be upset with me if I drank alcohol, smoked cigarettes, or used other drugs," "… told me stories of others who have gotten in trouble because of drinking alcohol, smoking cigarettes, or using other drugs," and "… warned me about the dangers of drinking alcohol, smoking cigarettes, or using other drugs" (1 = never to 4 = many times).
Secrecy from friends	Laird, 2013	10-item Self-Concealment Scale in reference to	e.g. "I'm often afraid I'll reveal something to my best friend that I don't want to." Adolescents reported the extent to which agreed with

		secrecy from best friends	each statement using a five-point scale (1 = strongly disagree to 5 = strongly agree).
Sexting	Ojeda, 2019	Not a validated scale	Four items were used to evaluate sexting, each with five response options (0=Never, 1=Rarely, 2=Occasionally, 3=Often, and 4=Frequently). The questions assessed were sending ("I have sent erotic-sexual videos, images or messages to my boyfriend/girlfriend"), receiving ("I have received erotic-sexual videos, images or messages from my boyfriend/girlfriend"), third-party forwarding ("I have forwarded or shared erotic-sexual videos, images or messages of other boys or girls"), and receiving via an intermediary ("Someone sent me erotic-sexual videos, images or messages of other boys or girls").
Teasing			
Appearance conversation with friends	Webb, 2017	6-item Appearance Conversations with Friends Scale	A sample item is: "My friends and I talk about looking good through fitness and sport" (1 = never, 5 = very often). Averaging items formed the total score, with a possible range of 1–5.
Overt victimization (teasing)	Borschmann, 2020	Selected items from Gatehouse Bullying Scale	See full scale in Hamburger_2011_Measuring bullying victimization perpetration and.pdf
Peer teasing	Helfert, 2011	Peer teasing subscale of the Appearance-related Social Pressure Questionnaire (FASD)	Composed of direct and intended kinds of verbal and non-verbal provocations, like disparaging comments and gestures by peers, e.g., "Other teenagers give me nicknames because of my body shape."
Appearance teasing	Webb, 2017	Weight Teasing subscale of the Perceptions of Teasing Scale	Sample items are: "Do your parents tease you about your weight or looks?" and "Do people your age call you names because of your appearance?" Participants rated the frequency of the behavior (1 = never, 5 = always), and the degree to which the behaviour was upsetting (1 = not upset, 5 = very upset).

Thematic category: Perpetration (e.g. bullying, aggression)

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Espelage, 2012, USA	Observational: prospective cohort study (2 timepoints with 6-month interval)	1391 students from 4 Midwestern middle schools, aged 10-15 years (mean age: 13.9±1.05 years)	Bullying perpetration in the past 30 days, measured at T1 (Spring 2008) (9-item Illinois Bully Scale, self-reported)	Outcomes measured at T2 (Fall 2008) using the modified AAUW sexual harassment survey: - Sexual harassment perpetration (sexual comments, spreading rumours, pulling at clothing) in past year

				- Forced sexual contact perpetration (forcing someone to kiss/do something sexual besides kissing/touching your private parts) in past year Statistics: multivariate regression modelling
Farrington, 2011, UK	Observational: prospective cohort study (Cambridge Study in Delinquent Development; interviews and tests at age 8, 10, 14, 16 and 18)	411 boys from South London followed up from age 8 to age 18	Bullying perpetration at age 14 (self-reported)	Outcome assessed at age 15-18: - Violence (self-reported) Statistics: logistic regression modelling
Foshee, 2014, USA	Observational: prospective cohort study (2 timepoints with 2-year interval)	1154 adolescents in grade 6 (at T1) and grade 8 (at T2) (mean age not available)	Assessed at T1 (measures created from the Nonphysical Aggression and Physical Aggression subscales of the Problem Behavior Frequency Index, self-reported): - Direct bullying perpetration - Indirect bullying perpetration	Outcomes measured at T2: - Physical dating violence perpetration: 1 question ("How many times have you ever used physical force against someone you were dating or on a date with (such as hitting, pushing, shoving, kicking, or assaulting them with a weapon) that was not in self-defense or play?"; response 0 to 10 times or more) Statistics: multivariate logistic regression modelling
Foshee, 2016, USA	Observational: prospective cohort study (2 timepoints with 6-month interval)	3161 dating adolescents attending the 8th (aged 13-14 years), 9th (aged 14-15) and 10th (aged 15-16) grade at T1 (Fall) and T2 (next Spring) in 19 rural schools	Bullying perpetration at T1 (questionnaire; self- and peer-reported)	Outcomes measured at T2: - Physical dating violence perpetration: short version of the Safe Dates Physical Dating Abuse Perpetration Scale; adolescents were asked how often they had perpetrated six physically violent acts against someone they were dating or on a date with during the previous 3 months (not counting acts committed in self-defense or play), i.e. "pushed, grabbed, shoved, or kicked," "slapped or scratched," "physically twisted their arm," "hit them with a fist or something else

				hard," "beat them up," and "assaulted them with a knife or gun." Response categories ranged from zero (0) to 10 times or more (5). A follow-up dating violence perpetration score was created by summing and averaging the responses. Statistics: conditional process modelling using path analytic regression and bootstrapping
Hemphill, 2011, Australia	Observational: prospective cohort study (Victorian cohort of the International Youth Development Study, 3 timepoints with 3-year interval between T1 and T2 and 1-year interval between T2 and T3)	701 students aged between 11.9 and 14.4 years (mean age: 12.9±0.4 years) at T1	Bullying perpetration at T1 and T2 (modified version of the Communities that Care survey, self-reported)	Outcomes measured at T3: - Theft: single item: stealing something worth more than $10. Scores were dichotomised; students were classified as reporting no involvement in these behaviours in the past 12 months (score of 0) or having engaged in these behaviours at least 1–2 times in the past 12 months (score of 1). - Violent behaviour: 3-item scale (e.g. how many times in the past year have you beat up someone so badly that they probably needed to see a doctor or nurse?) with response options ranging from 1 (Never) to 8 (40 or more times). Scores were dichotomised (0 = no involvement in violent behaviour, 1 = having engaged in violent behaviour at least 1–2 times). - Binge drinking: single item: "How many times in the last 2 weeks they have had five or more drinks in a row?", rated from 1 (none) to 6 (10 or more times). Scores were dichotomised; students were classified as reporting no involvement in these behaviours in the past 12 months (score of 0) or having engaged in

				these behaviours at least 1–2 times in the past 12 months (score of 1). - Marijuana use: single item: "In the past 30 days on how many occasions have you used marijuana (pot, weed, grass)?". Scores were dichotomised; students were classified as reporting no involvement in these behaviours in the past 12 months (score of 0) or having engaged in these behaviours at least 1–2 times in the past 12 months (score of 1). - Depressive symptoms: self-report Short Mood and Feelings Questionnaire. Respondents with a score of 11 were coded 1 (displaying depressive symptoms); all other respondents were coded 0 (not showing depressive symptoms). Statistics: logistic regression analysis
Ingram, 2020, USA	Observational: prospective cohort study (5 waves with 6-month intervals)	894 students in the 5-7th grade (at baseline) attending 4 Midwestern middle schools	Peer bullying perpetration ('peer aggression') at waves 3 and 4 (9-item Illinois Bully Scale, self-reported)	Outcomes measured at wave 5: - Cannabis use: 7-point Likert-type scale, number of days in the past 30 days they used cannabis - Depressive symptoms: 6-item Orpinas Modified Depression Scale - Delinquency: 8-item scale based on Jessor and Jessor's General Deviant Behavior Scale (self-reported) Statistics: latent class analysis, followed by mixture modelling with full information likelihood estimation
Kaltiala-Heino, 2010, Finland	Observational: prospective cohort study (Adolescent Mental Health Cohort Study; 2 timepoints with 2-year interval)	2070 students aged 15 years (mean age: 15.5±0.4 years) at T1 and 17 at T2	Bullying perpetration at T1 (1 question derived from the WHO youth health study, self-reported)	Outcomes measured at T2: - Depression: R-BDI, a Finnish modification of the 13-item Beck Depression Inventory Statistics:

				logistic regression analyses
Kendrick, 2012, Canada	Observational: prospective cohort study (2 timepoints, approximately 1-year interval)	880 adolescents, aged 12-16 years (mean age: 13.72±0.78 years) at T1, attending 7 junior high schools in central Sweden	Bullying perpetration at T1 (measure developed by Alsaker and Brunner 1999, self-reported)	Outcomes measured at T2: - Depression: adapted from the Center for Epidemiologic Studies Depression Scale - Committing property crimes (=theft, burglary, shoplifting, and motor vehicle theft) : 8-item scale created for this study that asks how often youths committed property crimes during the past year. Responses were recorded on a five-point scale, with possible responses being: (1) 'No, it has not happened'; (2) 'Once'; (3) '2–3 times'; (4) '4–10 times'; and (5) 'More than 10 times'. Statistics: cross-lagged regression modelling
Klomek, 2008, USA	Observational: prospective cohort study (2 timepoints with 10-year interval)	2348 boys, born in 1981 in Finland, aged 8 at T1 and 18 at T2	Bullying perpetration at T1 (questionnaires filled in by parents, teachers and children)	Outcomes measured at T2: - Suicidal ideation: 9 items of the BDI (self-reported) Statistics: logistic regression modelling
Lösel, 2011, UK	Observational: prospective cohort study (3 timepoints with a 2.4-year interval between T1 and T2 and a 4.7-year interval between T2 and T3)	T1: 637 German children with a mean age of 6.6 years ± 9.9 months T2: 557 of these children, mean age of 9 years ± 10.6 months T3: 596 of these children, mean age of 13.7 years ± 11.2 months	School bullying perpetration at T2 (selected items from the Olweus Bully/Victim questionnaire, self-reported)	Outcomes measured at T3: - Emotional problems: Strengths and Difficulties Questionnaire (SDQ); self-reported - Anxious/depressive problems: SPS; self-reported - Anxious/depressive problems: Social Behaviour Questionnaire (SBQ); mother-reported - Self-esteem problems: SPS; self-reported - Delinquency problems: SBQ - Delinquency problems: German Delinquency self-report scale (DBS) - Violent offences: DBS - Physical aggression: SBQ

				Statistics: hierarchical regression analysis controlled for risk factors at T1
Luukkonen, 2009, Finland	Observational: case-control study	508 adolescents aged 12-17 years admitted to a psychiatric hospital (suicide attempters: 26 boys and 78 girls, non-suicidals: 161 boys and 148 girls)	Bullying perpetration (K-SADS-PL; self-reported, complemented by parent-reported in case of missing or unreliable data)	Outcome: - Suicide attempts: K-SADS-PL; at least 1 suicide attempt with definite suicidal intent or if it was life-threatening (for example, the adolescent was briefly unconscious) Statistics: forward step-wise logistic regression analysis
McVie, 2014, UK	Observational: prospective cohort study (6 timepoints with 1-year intervals)	Around 4300 children, starting secondary school at around age 12 (T1), surveyed annually until age 17 (T6)	High bullying perpetration between T2 and T5 (age 13 and 16) (adapted from the Olweus Bully/Victim questionnaire, self-reported)	Outcomes measured at T6 (age 17): - Self-reported involvement in violence: asked whether they had committed any of the following five acts of violence during the previous year: "hit or picked on someone because of their race or skin color"; "hit, kicked, punched or attacked someone with the intention of really hurting them"; "stolen money or property that someone was holding, carrying or wearing using threats or actual force or violence"; "hurt or injured any animals or birds on purpose"; and "carried a knife or other weapon for protection or in case it was needed in a fight." A binary measure was created indicating whether or not the respondent had reported any violence at age 17. Statistics: logistic regression modelling
Moore, 2014, Australia	Observational: prospective cohort study (Raine birth cohort with measurements at 5, 10, 14 and 17 years of age)	1754 17-year olds, of whom 1590 completed the questionnaire at age 14	Bullying ('peer aggression') perpetration at age 14 (self-reported questionnaire designed for this study)	Outcomes measured at age 17: - Depression: Beck Depression Inventory for Youth (BDI-Y) - Externalizing problems (aggressive/delinquent): Youth Self Report (YSR/11-18)

				- Internalizing problems (withdrawn, somatic complaints, anxious/depressed): Youth Self Report (YSR/11-18) - Harmful alcohol use: self-reported questionnaire designed for this study - Harmful cannabis use: self-reported questionnaire designed for this study Statistics: multivariate logistic regression modelling
Pellegrini, 2001, USA	Observational: prospective cohort study (2 timepoints with 1-year interval)	129 students from 2 middle schools in rural North America entering grade 6 (mean age 12.8 years, SD not available) at T1 and ending grade 7 (mean age 14.01 years, SD not available) at T2	Bullying perpetration at T1 (Olweus' self-report The Senior Bully-Victim Questionnaire)	Outcomes measured at T2: - Sexual harassment perpetration: 12 questions with a 1 (never) to 5 (daily) response scale, addressing issues of sexual commenting, touching, forcing sexual contact, completed for each of the focal youngsters by each of the research associates who spent time in that school interviewing and observing target youngsters during the 2 school years. Each youngster was rated by seven researchers. Their scores were the mean of these seven ratings across all 12 items. The score, then, represents the average response/item. Statistics: mediational path modelling using hierarchic regression procedures
Pisarska, 2020, Poland	Observational: prospective cohort study (2 timepoints with a two-year interval)	551 adolescents, aged 16 years (grade 10) at T1 and 18 years (grade 12) at T2, attending public/non-public general, technical high schools and basic vocational schools	Cyberbullying perpetration at T1 (no information on measurement instrument or who reported the cyberbullying perpetration)	Outcomes measured at T2: - Gambling involvement that includes both gambling-related behaviours and problems (self-reported) Statistics: generalised linear models with gamma variation

Prinstein, 2004, USA	Observational: prospective cohort study (2 timepoints with 6-year interval)	148 girls in grades 4-6 at T1 and in grades 10-12 (aged 15-18 years, mean age 16.82±0.86 years) at T2	Peer aggression and/or disruptive behaviour at T1 (peer nominations)	Outcomes measured at T2: - Externalizing behaviour: Youth Self Report (YSR) + Child Behavior Checklist (CBCL) - Heavy episodic drinking: 2 items in the Survey of Risk Taking Behaviors - Marijuana use: 1 item in the Survey of Risk Taking Behaviors - Hard drug use: number of times in past year Statistics: hierarchical linear regression modelling
Stallard, 2013, UK	Observational: prospective cohort study (2 timepoints with 6-month interval)	3964 adolescents aged 12-16 years at T1 attending 8 secondary schools (mean age not available)	Bullying perpetration (single item, self-reported) at T1	Identified from included studies SR of Heerde 2018, which in its turn was identified from the reference list of Borschmann 2020 Outcomes measured at T2: - Self-harming thoughts and behaviours: 2 questions from the ALSPAC study ("Have you ever thought about hurting yourself, even if you would not really do it, in the last 6 months?", categorised as *not at all* vs *once or more;* and "Have you ever hurt yourself on purpose in any way (e.g. by taking an overdose of pills or by cutting yourself), in the last 6 months?", categorised as *not at all* vs *once or more)* Statistics: multivariate logistic regression modelling
Winsper, 2012, UK	Observational: prospective cohort study (2 timepoints with 2-year interval)	6043 children aged 8 at T1 and between 10.4 and 13.6 (mean 11.9 years, SD not available) at T2	Peer bullying perpetration at T1 (reported by mother and child)	Outcomes measured at T2: - Suicide ideation in past 2 years: 1 question ("Have you thought about killing yourself?") Statistics: logistic regression modelling

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
12-18 years				
Depressive symptoms (T3)	Bullying perpetration (T1)	Not statistically significant: ££ aOR: 0.97, 95%CI [0.58;1.63] ¥ (p>0.05)	1, 687	Hemphill, 2011
	Bullying perpetration (T2)	Not statistically significant: ££ aOR: 1.39, 95%CI [0.85;2.27] ¥ (p>0.05)	1, 701	
Depression (T2)	Bullying perpetration (T1)	*Girls* Not statistically significant: 3/15 vs 117/1148 § aOR: 1.6, 95%CI [0.4;7.4] ¥ (p>0.05)	1, 15 vs 1148	Kaltiala-Heino, 2010
		Boys Statistically significant: 8/39 vs 49/855 § aOR: 3.1, 95%CI [1.2;7.7] (p=0.02) *with harm for bullying perpetration*	1, 39 vs 855	
Depression (T2) (β±SE)	Bullying perpetration (T1)	Not statistically significant: -0.01±0.03 £† (p=0.869)	1, 880	Kendrick, 2012
Depression (age 17)	Peer bullying perpetration vs no perpetration (age 14)	Statistically significant: 57/206 vs 138/750 § aOR: 1.81, 95%CI [1.19;2.76] (p<0.05) *with harm for peer bullying perpetration*	1, 206 vs 750	Moore, 2014
Internalizing problems (age 17)		Not statistically significant: 38/129 vs 159/839 § aOR: 1.53, 95%CI [0.93;2.52] ¥ (p>0.05)	1, 129 vs 839	
Self-harming thoughts and behaviours (T2)	Regular bullying perpetration vs never bullying (T1)	*Boys* Not statistically significant: 34/290 vs 88/1379 § aOR: 1.29, 95%CI [0.79;2.10] ¥ (p>0.05)	1, 290 vs 1379	Stallard, 2013
		Girls Not statistically significant: 18/148 vs 149/1350 § aOR: 0.59, 95%CI [0.31;1.10] ¥ (p>0.05)	1, 148 vs 1350	
Suicide attempts	Bullying perpetration	*Boys* Not statistically significant: 6/42 vs 7/91 § aOR: 2.16, 95%CI [0.59;7.87] ¥ (p=0.244)	1, 42 vs 91	Luukkonen, 2009
		Girls Statistically significant: 10/30 vs 27/141 § aOR: 3.27, 95%CI [1.08;9.95] (p=0.037) *with harm for bullying perpetration*	1, 30 vs 141	
Binge drinking (T3)	Bullying perpetration (T1)	Not statistically significant: ££ aOR: 1.13, 95%CI [0.76;1.68] ¥ (p>0.05)	1, 687	Hemphill, 2011
	Bullying perpetration (T2)	Statistically significant: ££ aOR: 1.67, 95%CI [1.12;2.50]	1, 701	

		(p<0.05) *with harm for bullying perpetration at T2*		
Harmful alcohol use (age 17)	Peer bullying perpetration vs no perpetration (age 14)	Statistically significant: 102/390 vs 93/566 § aOR: 1.76, 95%CI [1.23;2.53] (p<0.05) *with harm for peer bullying perpetration*	1, 390 vs 566	Moore, 2014
Harmful cannabis use (age 17)		Not statistically significant: 29/98 vs 162/861 § aOR: 1.61, 95%CI [0.96;2.70] ¥ (p>0.05)	1, 98 vs 861	
Marijuana use (T3)	Bullying perpetration (T1)	Not statistically significant: ££ aOR: 1.48, 95%CI [0.76;2.89] ¥ (p>0.05)	1, 687	Hemphill, 2011
	Bullying perpetration (T2)	Not statistically significant: ££ aOR: 1.42, 95%CI [0.76;2.64] ¥ (p>0.05)	1, 701	
Gambling involvement (T2)(B)	Cyberbullying perpetration (T1)	Statistically significant: 0.016 £ (p=0.010) *with harm for cyberbullying perpetration*	1, 511	Pisarska, 2020
Externalizing problems (age 17)	Peer bullying perpetration vs no perpetration (age 14)	Not statistically significant: 55/170 vs 142/798 § aOR: 1.12, 95%CI [0.68;1.84] ¥ (p>0.05)	1, 170 vs 798	Moore, 2014
Theft (T3)	Bullying perpetration (T1)	Not statistically significant: ££ aOR: 1.21, 95%CI [0.64;2.31] ¥ (p>0.05)	1, 687	Hemphill, 2011
	Bullying perpetration (T2)	Statistically significant: ££ aOR: 2.21, 95%CI [1.27;3.85] (p<0.01) *with harm for bullying perpetration at T2*	1, 701	
Property crimes (T2)(β±SE)	Bullying perpetration (T1)	Not statistically significant: 0.11±0.06 £† (p=0.082)	1, 880	Kendrick, 2012
Involvement in violence (T6)(β±SE)	High bullying perpetration (T2-T5)	Statistically significant: 0.93±0.13 aOR: 2.54, 95%CI [1.96;3.30] (p<0.001) *with harm for high bullying perpetration*	1, 2292	McVie, 2014
Violent behaviour (T3)	Bullying perpetration (T1)	Not statistically significant: ££ aOR: 1.17, 95%CI [0.71;1.93] ¥ (p>0.05)	1, 687	Hemphill, 2011
	Bullying perpetration (T2)	Statistically significant: ££ aOR: 2.21, 95%CI [1.27;3.85] (p<0.01) *with harm for bullying perpetration at T2*	1, 701	
Self-reported violence (age 15-18)	Bullying perpetration (age 14)	*Controlled for 20 explanatory risk factors:* Not statistically significant: ££ Partial OR: 2.02, 95%CI [0.97;	1, 411	Farrington, 2011

		£££££] † (p=0.059) *Controlled for behavioural risk factor troublesomeness:* Not statistically significant: ££ Partial OR: 1.82, 95%CI [1.00; £££££] † (p=0.050) *Controlled for behavioural risk factor anti-social personality:* Not statistically significant: ££ Partial OR: 1.78, 95%CI [0.98; £££££] † (p=0.058) *Controlled for behavioural risk factor teacher-rated aggressiveness:* Statistically significant: ££ Partial OR: 1.91, 95%CI [1.02; £££££] (p=0.033) *with harm for bullying perpetration*		
Sexual harassment (T2)	Bullying perpetration (T1)	Not statistically significant: ££££† (p=0.15)	1, 129 §	Pellegrini, 2001
Physical dating violence perpetration (T2) (β)	Bullying perpetration without bullying victimization (T1)	Statistically significant: 0.0303 £ (p=0.0296) *with harm for bullying perpetration*	1, number of participants not reported †	Foshee, 2016
5-18 years				
Depressive symptoms (W5) (mean±SE)	Peer bullying perpetration vs low aggression (W3-4)	Not statistically significant: 1.00±0.09 vs 0.88±0.43 $ MD: 0.12, 95%CI [0.07;0.17] (p>0.05)	1, 98 vs 361	Ingram, 2020
Self-reported anxious/depressive problems (T3) (adjusted correlation coefficient (r_{adj}))	School bullying perpetration (T2)	Statistically significant: 0.11 £ (p<0.05) *with harm for school bullying perpetration*	1, at least 486 †	Lösel, 2011
Mother-reported anxiousness/social withdrawal (T3) (r_{adj})		Not statistically significant: 0.00 £† (p>0.05)	1, at least 486 †	
Emotional problems (T3) (r_{adj})		Not statistically significant: 0.03 £† (p>0.05)	1, at least 486 †	
Self-esteem problems (T3) (r_{adj})		Not statistically significant: 0.05 £† (p>0.05)	1, at least 486 †	
Suicidal ideation (T2)	Frequent bullying perpetration vs never bullying (T1)	Not statistically significant: 48/172 vs 30/938 § £££† (p>0.05)	1, 172 vs 938	Klomek, 2008
Suicide ideation (T2)	Bullying perpetration at T1 (child-reported)	Statistically significant: ££ aOR: 3.60, 95%CI [1.46;8.84] (p<0.05)	1, 4404	Winsper, 2012

		with harm for bullying perpetration		
	Bullying perpetration at T1 (mother-reported)	Not statistically significant: ££ aOR: 1.25, 95%CI [0.77;2.02] ¥ (p>0.05)	1, 4990	
Heavy episodic drinking (T2) (β)	Aggression and/or disruptive behaviour (T1)	Not statistically significant: 0.15 £† (p>0.05)	1, 148 §	Prinstein, 2004
Hard drug use (T2)(β)		Not statistically significant: -0.02 £† (p>0.05)		
Marijuana use (T2) (β)		Not statistically significant: 0.10 £† (p>0.05)		
Cannabis use (W5) (mean±SE)	Peer bullying perpetration vs low aggression (W3-4)	Statistically significant: 1.36±0.32 vs 0.50±0.11 $ MD: 0.86, 95%CI [0.80;0.92] (p<0.05) *with harm for peer bullying perpetration*	1, 98 vs 361	Ingram, 2020
Sexual harassment perpetration (T2) (B±SE)	Bullying perpetration (T1)	Statistically significant: 0.02±0.02 (p<0.001) *with harm for bullying perpetration*	1, 1391	Espelage, 2012
Forced sexual contact perpetration (T2) (B±SE)		Not statistically significant: 0.02±0.02 £† (p>0.05)		
Physical dating violence perpetration (T2)	Direct bullying perpetration (T1)	Statistically significant: ££ aOR: 1.36, 95%CI [1.11;1.68] (p<0.01) *with harm for direct bullying perpetration*	1, 1154	Foshee, 2014
	Indirect bullying perpetration (T1)	Not statistically significant: ££ aOR: 1.26, 95%CI [0.93;1.73] ¥ (p>0.05)		
Mother-reported physical aggression (T3) (r_adj)	School bullying perpetration (T2)	Statistically significant: 0.14 £ (p<0.001) *with harm for school bullying perpetration*	1, at least 486 †	Lösel, 2011
Self-reported violent offences (T3) (r_adj)		Statistically significant: 0.21 £ (p<0.001) *with harm for school bullying perpetration*		
Self-reported delinquency (T3) (r_adj)		Statistically significant: 0.18 £ (p<0.001) *with harm for school bullying perpetration*		
Mother-reported delinquency (T3) (r_adj)		Statistically significant: 0.11 £ (p<0.01) *with harm for school bullying perpetration*		
Delinquency (W5) (mean±SE)	Peer bullying perpetration vs low aggression (W3-4)	Statistically significant: 0.42±0.06 vs 0.24±0.29 $ MD: 0.18, 95%CI [0.15;0.21] (p<0.05)	1, 98 vs 361	Ingram, 2020

		with harm for peer bullying perpetration		
Parent-reported externalizing behaviour (T2) (β)	Aggression and/or disruptive behaviour (T1)	Not statistically significant: 0.13 £† (p>0.05)	1, 148 §	Prinstein, 2004
Adolescent-reported externalizing behaviour (T2) (β)		Not statistically significant: 0.20 £† (p>0.05)		

Mean±SD or B±SE (unless indicated otherwise), MD: mean difference, OR: odds ratio, aOR: adjusted odds ratio, SD: standard deviation, SE: standard error, B: unstandardized coefficient, β: standardized coefficient

$ The outcome measures and effect measures represent the risk factor/exposure, not the outcome

£ No SE and/or CI available; or no information on magnitude of effect available to assess variability of results.

££ No raw data available

£££ No adjusted OR reported

££££ No raw data or effect size reported

£££££ Upper limit of 95%CI not reported

¥ Imprecision (large variability of results)

† Imprecision (lack of data)

§ Imprecision (limited sample size or low number of events)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Espelage, 2012	Unclear, no information on how the 4 middle schools were selected for this longitudinal study	Yes, sexual harassment and forced sexual contact were assessed via a non-validated scale	No, analyses were adjusted for age, gender and sexual harassment and forced sexual contact at T1	Unclear, no attrition rates mentioned in this study	No
Farrington, 2011	Yes, the sample of the Cambridge Study in Delinquent Development was chosen by taking all the boys who were then aged 8–9 and on the registers of six state primary schools within a one mile radius of a research office that had been established.	Yes, bullying was assessed via a non-validated scale. In addition, it is unclear how self-reported violence was assessed.	Unclear, analyses were controlled for 20 explanatory predictors (daring, hyperactivity, psychomotor clumsiness, non-verbal IQ, verbal IQ, attainment, extraversion, neuroticism, height, weight, popularity, parent conviction, sibling delinquency, young mothers, poor child-rearing, disrupted family, family income, family size, housing, social class) and 3 behavioural predictors (troublesomeness, anti-social personality, teacher-rated	Unclear, at age 18 95% of all boys were interviewed, but no analyses were done to investigate if the 5% that dropped out differed from the 95% that did not	No

			aggressiveness). However, it is unclear if and why these 3 behavioural predictors were taken into account separately, and not in combination with the explanatory predictors.		
Foshee, 2014	Yes, the study selected 3 public school systems in 3 primarily rural North Carolina counties. It is however unclear how the schools were selected. In addition, to control the temporality of relationships, adolescents who reported having ever hit or threatened a dating partner at the 6th-grade assessment were eliminated from the analyses.	Yes, both bullying and physical dating violence perpetration were assessed via a non-validated scale. In addition, the use of only two items to assess each type of bullying and the lack of items tapping into cyberbullying could have resulted in underreporting of bullying and potential misclassification of bullying status.	Yes, although analyses were controlled for parent education, family structure, family conflict, sex and race/ethnicity, it is likely that there is residual confounding (e.g. suicide ideation at baseline)	Unclear, attrition rates are 25%, but no analyses were done to investigate if the drop-outs out differed from the remaining sample	No
Foshee, 2016	Unclear, the study selected 3 school systems in 3 primarily rural North Carolina counties. It is however unclear how the schools were selected (even in the original study by Ennett 2008).	Yes, both bullying and physical dating violence perpetration were assessed via a non-validated scale.	No, models controlled for sex, race of the adolescent, family structure, parent education, grade, family aggression and lifetime dating violence perpetration at T1. The authors themselves indicate that residual confounding may have occurred (lack of empathy, need for power and control, narcissism, lack of social competence and problem solving skills, and exposure to poor parenting, going to schools that tolerate violence and living in neighbourhoods	Yes, in the complete (sample without missing data) versus the full sample of daters (including youth with missing data), there were significantly fewer boys, more white youth, fewer black youth, and fewer single parent households. However, there were no differences in the two samples in the percentages of adolescents of race/ethnicities other than white or black, parent education, family aggression,	No

			that provide models of aggressive behaviour).	bullying perpetration, bullying victimization, the four mediators, and baseline and follow-up dating violence.	
Hemphill, 2011	No, a two-stage cluster sampling approach was utilized to recruit students into the International Youth Development Study	Yes, both bullying perpetration and the outcomes were assessed via a non-validated scale and were self-reported	Yes, although analyses were controlled for gender, student impulsivity, concentration/attention deficits, adolescent interaction with antisocial friends, family history of antisocial behaviour, poor family management, family conflict, academic failure, and – in the case of binge drinking and marijuana use – parental attitudes favourable to drug use, there may be residual confounding (e.g. outcomes at T1)	Unclear, no information provided	No
Ingram, 2020	Unclear, the study piggybacked on another larger study that selected 4 middle schools. The authors do not mention this larger study or how the schools were selected.	Yes, delinquency and cannabis use were self-reported	Yes, family violence, poly-victimization, identity components (gender, sexuality, disability etc) and socioeconomic status were not taken into account	Unclear, the authors indicate that missing data ranged from 4-25% on variables across the study period, but provide no further explanation or analyses if the drop-outs differed from the remaining sample.	No
Kaltiala-Heino, 2010	No, the Youth and Mental Health Study is a population-based, representative (cluster sampling), prospective cohort study	Yes, bullying was assessed via a non-validated scale. Moreover, it was self-reported. The effect of social desirability might result in underreporting.	No, analyses were adjusted for depression at T1, parental education and family structure	Unclear, the authors themselves indicate that "among drop-outs, psychosocial and health problems may be more common than among participants. Thus, their absence could result in the presented	Yes, no correction for multiple testing was performed, because of the exploratory nature of the analyses

				prevalence rates being underestimates at T1. The response rate in the Adolescent Mental Health Cohort follow-up was satisfactory. The attrition in the Adolescent Mental Health Survey follow-up was, however, associated with faring worse in the baseline survey. It is likely that in the follow-up, the prevalence of psychosocial and mental health problems were to some extent an underestimate." However, they do not report any analyses on this matter.	
Kendrick, 2012	Unclear, the sample included participants from an ongoing longitudinal study within seven junior high schools (grades 7 through 9) in a town in central Sweden. The authors do not mention this larger study or how the schools were selected.	Yes, both the exposure and outcomes were self-reported	No, a design was used that controlled both for stability over time and for cross-sectional intercorrelations of all variables	No, 108 of 988 participants did not have data for both timepoints. Only students who participated at both timepoints were included. The group of 108 that was removed did not differ significantly from the participants in the analytical sample.	No
Klomek, 2008	No, the study included a representative sample of 10% of all Finnish children born in 1981	Yes, bullying was assessed via a non-validated scale	Yes, although the analyses accounted for depressive symptom experiencing, it is likely that there is residual confounding	Unclear, the authors report that attrition rates after 10 years were 20.3%, but provided no further explanation or analyses if the drop-outs differed from the remaining sample	No
Lösel, 2011	Yes, no random sampling. The families were	Yes, bullying perpetration was only	Yes, although analyses were controlled for	Unclear, the authors state in their tables that	The Erlangen-Nuremberg Development

	recruited through church and municipal kindergartens. Participation depended on the willingness of the kindergarten leaders.	measured by a few items, the assessment was carried out in an interview what may have led to more influences of response tendencies and self-presentation than in anonymous questionnaire surveys, and mothers' reports may have limited validity with regard to the less visible internalizing problems	family (e.g. mother-reported critical life events, socioeconomic status, punishment) and individual risk factors (mother-reported emotional problems and conduct disorder) that were measured at T1 (more than two years before bullying perpetration), other factors (e.g. self-reported emotional and conduct problems) were not taken into account	"n varies due to missing data", but do not provide further information	and Prevention Project is a combined experimental and prospective longitudinal study. In the subgroup that received the programme (=a parent training on positive parenting and a child social skills training), the T1 assessment of the control variables was carried out more than one year after the intervention. Therefore, the authors thought it was not necessary to restrict their current analyses to the untreated children ("Because we assessed the risk factors long before bullying and victimization we also avoided confound measurement, i.e. that the risk-controlled predictions only addressed change scores").
Luukkonen, 2009	No, no indication. The study sample comprised all adolescent patients hospitalized in the geographically	Unclear, it is not clear if the K-SADS-PL is adequate to study bullying perpetration	Yes, the authors themselves state that they were not able to assess personality features such as impulsivity, aggressiveness and identity	No, no indication	No

	large area of northern Finland		confusion, which may confound the relationships found in the study		
McVie, 2014	No, a census approach was used, with all secondary schools in Edinburgh being invited to participate and all parents being asked to consent to their children taking part. The final cohort consisted of around 92% of the total population of young people who were enrolled to start secondary school, at around the age of 12, in 1998.	Yes, both bullying perpetration and involvement in violence perpetration were assessed via a non-validated scale.	No, analyses were adjusted for resilience measures at the individual level (gender, school attainment, positive attributes and prosocial behaviour, impulsivity, social alienation, self-esteem), family level (family structure, socioeconomic status, eligibility for free school meals, parental supervision and monitoring, parent-child conflict, parental interest in education) and community level (economic deprivation, neighbourhood stability, neighbourhood crime rate)	Unclear, the authors report that response rates ranged from 96% at age 12 to 81% at age 17, but provided no further explanation or analyses if the drop-outs differed from the remaining sample	No
Moore, 2014	No, no indication	Yes, peer aggression, harmful alcohol use and harmful cannabis use were assessed using a questionnaire specifically designed for the study. The questionnaire did not collect information about the severity or type of peer aggression and did not assess cyberaggression independently of traditional forms. In addition, dichotomization of the outcome variables may	No, analyses were adjusted for sex, family income, mothers mental health, family structure, substance use at 14 years, and pre-existing mental health at 5 and 14 years	Yes, the authors themselves indicate that selective attrition among socially disadvantaged families may have occurred	No

		have resulted in a loss of information.			
Pellegrini, 2001	Unclear, participants were sampled from the total population, i.e. two middle schools serving the whole county. However, no further explanation on the sampling method.	Yes, the authors themselves indicate that the observational techniques used were not sensitive enough to record the youngster's many verbalizations	Yes, although the mediational path analysis controlled for the indirect path (from bullying to dating frequency at the end of 6th grade to sexual harassment) using hierarchic regression procedures before entering the direct path, other confounding factors (e.g. data from primary school) may still be present	No, the attrition rate was limited to 6.5%. Moreover, the authors mention that "the youngsters who left the study did not seem to differ significantly from those who remained in terms of variables used in this study".	No
Pisarska, 2020	No, random sampling of clusters (classrooms)	Yes, exposure and outcome measures were self-reported. These data are burdened with recall and social desirability bias.	No, analyses are adjusted to control for demographics and T1 gambling involvement	Yes, only 65% response rate from T1 to T2. Youth omitted because of missing data may have on average a higher Gambling Involvement Index score. This selection of students in longitudinal analyses may have biased the sample towards lower risk students in gambling involvement.	No
Prinstein, 2004	Unclear, participants attended three elementary schools. It is unclear how these schools were selected for this longitudinal study.	Yes, aggression and/or disruptive behaviour was assessed via a non-validated scale (peer nominations)	Yes, hierarchical linear regression modelling accounted for age, ethnicity and social preference, but did not take into account baseline externalizing behaviour, hard drug or marijuana use, or heavy episodic drinking	Unclear, the authors indicate that " the collection of data at multiple time points between childhood and adolescence may […] help to reduce attrition rates in long-term longitudinal studies". However, they do not mention attrition rates in this study.	No
Stallard, 2013	Unclear, the study piggybacked on an RCT that selected 8	Yes, exposure and outcome measures were self-reported. Moreover, only	No, analyses accounted for baseline age, trial arm (although there was no	No, the authors showed that non-availability of self-harm data	The study was not specifically designed as a prospective

	schools. The authors do not mention how they obtained the sample (neither in the original RCT).	a subset of the self-harm questions from the ALSPAC study were used.	evidence that the trial intervention had any effect on either depression or self-harm behaviour at follow-up), gender, family structure, socioeconomic status, mood, peer and school relationships and support, alcohol/cannabis/drug use. Moreover, analyses were confined to the people who did not self-harm at T1.	was associated with: older year group (Year 11); who the participant lived with; high SMFQ score; alcohol and cannabis use; bullying others; family not owning a car; and low school membership. These variables were all included within the multiple imputation process.	longitudinal survey and as such the assessment of possible self-harm predictors were limited.
Winsper, 2012	No, 13971 children of 14541 women residing in Avon while pregnant and with an expected delivery date between April 1 1991 and December 31 1992, were alive at 12 months formed the original birth cohort.	Yes, data regarding suicide ideation were obtained by self-report rather than by clinical Examination	Yes, analyses accounted for age, gender, abuse, maladaptive parenting, exposure to domestic violence, internalizing and externalizing problems, but not for baseline suicide ideation	Yes, the authors analysed if dropout was random or selective. Those lost to follow-up were more often boys, had higher internalizing and externalizing scores, were of ethnic minority, had low birth weight, were born to single mothers, had lower educational level, lived in rental properties, and had parents engaged in manual labor jobs. Those exposed to one or more family adversities were less likely to have suicide data, as were those exposed to domestic violence. Those exposed to physical or sexual abuse were more likely to have attended the interview. Generally, participants	No

52

				who dropped out were exposed to more deprivation than the remaining participants.	

Certainty of the body of evidence

<u>**12-18 years**</u>

	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes, low number of events, lack of data and/or large variability of the results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

<u>**5-18 years**</u>

Bullying perpetration	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Aggression and/or disruptive behaviour	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample size
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	<u>**12-18 years**</u> There is limited evidence **with harm** for peer bullying perpetration when it comes to **depression/depressive symptom experiencing, binge drinking, harmful alcohol use, gambling involvement, suicide attempts, physical dating violence perpetration, <u>externalizing problems, property crime, and general violent behaviour</u>**: • Depression/depressive symptom experiencing: It was shown in one study that bullying perpetration resulted in a statistically significant increase in depression (Moore 2014). In another study, this statistically significant increase in depression was found in boys,

but not in girls (Kaltiala-Heino 2010).
However, in two other studies, a statistically significant increased risk of depression or depressive symptom experiencing in presence of earlier bullying perpetration could not be demonstrated (Hemphill 2011, Kendrick 2012).

- Binge drinking & harmful alcohol use:
  It was shown that bullying perpetration at age 16, but not at age 13, resulted in a statistically significant increased risk of binge drinking at age 17 (Hemphill 2011).
  In addition, another study showed that bullying perpetration resulted in a statistically significant increase in harmful alcohol use (Moore 2014).

- Gambling involvement:
  It was shown that cyberbullying perpetration resulted in a statistically significant increased risk of gambling involvement (Pisarska 2020).

- Suicide attempts:
  In one study, it was shown that bullying perpetration resulted in a statistically significant increase in suicide attempts in girls, but not in boys (Luukkonen 2009).

- Physical dating violence perpetration:
  It was shown that bullying perpetration resulted in a statistically significant increased risk of physical dating violence perpetration (Foshee 2016).

- Externalizing problems, property crime, or general violent behaviour:
  It was shown in one study that involvement in high bullying perpetration resulted in a statistically significant increase in involvement in violence (McVie 2014). In another study, it was shown that bullying perpetration at age 16, but not at age 13, resulted in a statistically significant increased risk of theft and violent behaviour at age 17 (Hemphill 2011).
  However, a statistically significant increased risk of externalizing problems (Moore 2014), property crimes (Kendrick 2012) or self-reported violence (except when controlled for teacher-rated aggressiveness) (Farrington 2011) in the presence of earlier bullying perpetration could not be demonstrated.

There is limited evidence concerning the risk of **internalizing problems, self-harming thoughts and behaviour, (harmful) cannabis use or sexual harassment** in the presence of earlier bullying perpetration:

- Internalizing problems:
  A statistically significant increased risk of internalizing problems in presence of earlier bullying perpetration could not be demonstrated (Moore 2014).

- Self-harming thoughts and behaviour:
  A statistically significant increased risk of self-harming thoughts and behaviours in the presence of regular bullying perpetration could not be demonstrated (Stallard 2013).

- (Harmful) cannabis use:
  A statistically significant increased risk of (harmful) cannabis use in the presence of earlier bullying perpetration could not be demonstrated (Moore 2014, Hemphill 2011).

- Sexual harassment:
  A statistically significant increased risk of sexual harassment in the presence of earlier bullying perpetration could not be demonstrated (Pellegrini 2001).

Evidence is of very low certainty and results cannot be considered precise due to limited sample size, low number of events, large variability of results and/or lack of data.

5-18 years

1) Bullying perpetration

There is limited evidence **with harm** for peer bullying perpetration when it comes to **self-reported anxious/depressive problems, suicidal ideation, cannabis use, sexual harassment perpetration, physical dating violence perpetration, mother-reported physical aggression, self-reported violent offenses, or delinquency:**

- Self-reported anxious/depressive problems:
  It was shown that school bullying perpetration at the age of 9 years resulted in a statistically significant increased risk of self-reported anxious/depressive problems at 14 years of age (Lösel 2011).

- Suicidal ideation:
  It was shown that self-reported, but not mother-reported, bullying perpetration at the age of 8 years resulted in a statistically significant increased risk of suicidal ideation at 12 years of age (Winsper 2012). However, in another study, a statistically significant increased risk of suicidal ideation at 18 years of age in the presence of frequent bullying perpetration at the age of 8 years could not be demonstrated (Klomek 2008).

- Cannabis use:
  It was shown that school bullying perpetration in grade 5-7 students resulted in a statistically significant increased risk of cannabis use 6 months to one year later (Ingram 2020).

- Sexual harassment perpetration:
  It was shown that school bullying perpetration in children aged 10 to 15 years resulted in a statistically significant increased risk of sexual harassment perpetration 6 months later (Espelage 2012).

- Physical dating violence perpetration:
  It was shown that direct, but not indirect, bullying perpetration in grade 6 students resulted in a statistically significant increased risk of physical dating violence perpetration 2 years later (Foshee 2014).

- Mother-reported physical aggression:
  It was shown that school bullying perpetration at the age of 9 years resulted in a statistically significant increased risk of mother-reported physical aggression at 14 years of age (Lösel 2011).

- Self-reported violent offenses:
  It was shown that school bullying perpetration at the age of 9 years resulted in a statistically significant increased risk of self-reported violent offenses at 14 years of age (Lösel 2011).

- Delinquency:
  It was shown that school bullying perpetration at the age of 9 years resulted in a statistically significant increased risk of self-reported and mother-reported delinquency at 14 years of age (Lösel 2011). It was also shown that peer bullying perpetration in grade 5-7 students resulted in a statistically significant increased risk of self-reported delinquency 6 months to one year later (Ingram 2020).

There is limited evidence concerning the risk of **depressive symptoms, mother-reported anxiousness/social withdrawal, emotional problems, self-esteem problems, or forced sexual contact perpetration** in the presence of earlier bullying perpetration:

- Depressive symptoms:
  A statistically significant increased risk of depressive symptom experiencing in the presence of peer bullying perpetration 6 months to one year earlier, could not be demonstrated in grade 5-7 students (Ingram 2020).

- Mother-reported anxiousness/social withdrawal, self-reported emotional problems, or self-reported self-esteem problems:
  A statistically significant increased risk of mother-reported

anxiousness/social withdrawal, self-reported emotional problems or self-reported self-esteem problems at 14 years of age, in the presence of earlier school bullying perpetration at the age of 9 years, could not be demonstrated (Lösel 2011).

- Forced sexual contact perpetration:
  A statistically significant increased risk of forced sexual contact perpetration in the presence of bullying perpetration 6 months earlier in children aged 10 to 15 years, could not be demonstrated (Espelage 2012).

Evidence is of very low certainty.

2) Aggression and/or disruptive behaviour

There is limited evidence concerning the risk of **heavy episodic drinking, hard drug use, cannabis use, parent-reported externalizing behaviour or adolescent-reported externalizing behaviour** in the presence of earlier aggression and/or disruptive behaviour:
A statistically significant increased risk of heavy episodic drinking, hard drug use, cannabis use, parent-reported externalizing behaviour and adolescent-reported externalizing behaviour in adolescents aged 15 to 18 years, in the presence of aggression and/or disruptive behaviour 6 years earlier, could not be demonstrated (Prinstein 2004).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size.

Reference(s)	**Articles** Espelage DL, Basile KC, Hamburger ME. *Bullying perpetration and subsequent sexual violence perpetration among middle school students.* J Adolesc Health 2012, 50(1):60-5. doi: 10.1016/j.jadohealth.2011.07.015. Farrington DP, Ttofi MM. *Bullying as a predictor of offending, violence and later life outcomes.* Crim Behav Ment Health 2011, 21(2):90-8. doi: 10.1002/cbm.801. Foshee VA, McNaughton Reyes HL, Vivolo-Kantor AM, Basile KC, Chang L-Y, Faris R, Ennett ST. *Bullying as a longitudinal predictor of adolescent dating violence.* J Adolesc Health 2014, 55(3):439-44. doi: 10.1016/j.jadohealth.2014.03.004. Foshee VA, Benefield, TS, McNaughton Reyes HL, Eastman M, Vivolo-Kantor AM, Basile KC, Ennett ST, Faris R. *Examining explanations for the link between bullying perpetration and physical dating violence perpetration: Do they vary by bullying victimization?* Aggress Behav 2016,42(1):66-81. doi: 10.1002/ab.21606. Hemphill SA, Kotevski A, Herrenkohl TI, Bond L, Kim MJ, Toumbourou JW, Catalano RF. *Longitudinal consequences of adolescent bullying perpetration and victimisation: a study of students in Victoria, Australia.* Crim Behav Ment Health 2011, 21(2):107-16. doi: 10.1002/cbm.802. Ingram KM, Espelage DL, Davis JP, Merrin GJ. *Family Violence, Sibling, and Peer Aggression During Adolescence: Associations With Behavioral Health Outcomes.* Front Psychiatry 2020, 11:26. doi: 10.3389/fpsyt.2020.00026. Kaltiala-Heino R, Fröjd S, Marttunen M. *Involvement in bullying and depression in a 2-year follow-up in middle adolescence.* Eur Child Adolesc Psychiatry 2010, 19(1):45-55. doi: 10.1007/s00787-009-0039-2. Kendrick K, Jutengren G, Stattin H. *The protective role of supportive friends against bullying perpetration and victimization.* J Adolesc 2012, 35(4):1069-80. doi: 10.1016/j.adolescence.2012.02.014. Klomek AB, Sourander A, Kumpulainen K, Piha J, Tamminen T, Moilanen I, Almqvist F, Gould MS. *Childhood bullying as a risk for later depression and suicidal ideation among Finnish males.* J Affect Disord 2008, 109(1-2):47-55. doi: 10.1016/j.jad.2007.12.226. Lösel F, Bender D. *Emotional and antisocial outcomes of bullying and victimization at school: A follow-up from childhood to adolescence.* Journal of Aggression, Conflict and Peace Research 2011, 3(2), 89–96. doi:10.1108/17596591111132909. Luukkonen A-H, Räsänen P, Hakko H, Riala K, STUDY-70 Workgroup. *Bullying behavior is related to suicide attempts but not to self-mutilation among psychiatric inpatient adolescents.* Psychopathology 2009, 42(2):131-8. doi: 10.1159/000204764. McVie S. *The Impact of Bullying Perpetration and Victimization on Later Violence and Psychological Distress: A Study of Resilience Among a Scottish Youth Cohort.* Journal of School Violence 2014, 13:1, 39-58. doi: 10.1080/15388220.2013.841586

Moore SE, Norman RE, Sly PD, Whitehouse AJO, Zubrick SR, Scott J. *Adolescent peer aggression and its association with mental health and substance use in an Australian cohort.* J Adolesc 2014, 37(1):11-21. doi: 10.1016/j.adolescence.2013.10.006.
Pellegrini AD. *A longitudinal study of heterosexual relationships, aggression, and sexual harassment during the transition from primary school through middle school.* Journal of Applied Developmental Psychology 2001, 22(2):119-133. doi: 10.1016/S0193-3973(01)00072-7.
Pisarska A, Ostaszewski K. *Factors associated with youth gambling: longitudinal study among high school students.* Public Health 2020, 184:33-40. doi: 10.1016/j.puhe.2020.03.017
Prinstein MJ, La Greca AM. *Childhood peer rejection and aggression as predictors of adolescent girls' externalizing and health risk behaviors: a 6-year longitudinal study.* J Consult Clin Psychol 2004, 72(1):103-12. doi: 10.1037/0022-006X.72.1.103.
Stallard P, Spears M, Montgomery AA, Phillips R, Sayal K. *Self-harm in young adolescents (12-16 years): onset and short-term continuation in a community sample.* BMC Psychiatry 2013, 13:328. doi: 10.1186/1471-244X-13-328.
Winsper C, Lereya T, Zanarini M, Wolke D. *Involvement in bullying and suicide-related behavior at 11 years: a prospective birth cohort study.* J Am Acad Child Adolesc Psychiatry 2012, 51(3):271-282.e3. doi: 10.1016/j.jaac.2012.01.001.

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Bullying perpetration			
Peer bullying perpetration ('peer aggression')	Ingram, 2020 Espelage, 2012	9-item Illinois Bully Scale	Scale that consists of 9 items including those related to teasing and name-calling, social exclusion, and rumour spreading. Students were asked how often in the past 30 days they engaged in each behaviour (e.g., teased other students, excluded others from their group of friends, threatened to hit or hurt another student). Response options included 0 (Never), 1 (1 or 2 times), 2 (3 or 4 times), 3 (5 or 6 times), and 4 (7 or more times) on a 5-point Likert-type scale. [Ingram 2020: scores were averaged across Waves 3 and 4.]
School bullying perpetration	Lösel, 2011	Selected items from the Olweus Bully/Victim questionnaire	Three items addressed active involvement in beating/kicking, destroying others' property, and abusive language/picking on somebody. The response format at this age was simple: "never" (score 0), "sometimes" (1), and "often" (score 2) within a given time period (e.g. since Christmas). Sum scores were used as measures of bullying.
Bullying perpetration ('peer aggression')	Moore, 2014	Questionnaire designed for Raine cohort study	The questionnaire begins with the following statement, "Bullying is when someone is picked on by another person, or a group of people say nasty things to him or her. It is also when someone is hit, kicked, threatened, sent nasty notes or when no one talks to them." Although the Raine questionnaire used the term 'bullying' to describe these behaviours, 'peer aggression' is a better term as the definition provided to the participants does not refer to a power differential or repetition of the behaviour, two key components of bullying. 'Perpetrators of peer aggression' were those participants who endorsed having bullied other children at school.
Cyberbullying perpetration	Pisarska, 2020	No information on measurement instrument	/

Direct bullying perpetration	Foshee, 2014	Created from items in the Nonphysical Aggression and Physical Aggression subscales of the Problem Behavior Frequency Index	Adolescents were asked: "During the past 3 months, about how many times have you…,". The two direct bullying items were "picked on someone" and "hit or slapped another kid." Response categories ranged from 0 (0) to 10 times or more (5) in the past 3 months.
Indirect bullying perpetration			Adolescents were asked: "During the past 3 months, about how many times have you…,". The two indirect bullying items were "excluded another student from your group of friends," and "spread a false rumor about someone." Response categories ranged from 0 (0) to 10 times or more (5) in the past 3 months.
Bullying perpetration	Foshee, 2016	Questionnaire	Adolescents were instructed to write the first names (or initials) and ID number of up to five schoolmates who, in the past 3 months, "picked on or were mean" to them and up to five schoolmates whom they "picked on or were mean to." Students were instructed to disregard playful teasing and focus only on harmful actions. Social network analyses of these data were used to create bullying perpetration and bullying victimization variables. The bullying perpetration variable was based on self-reports of bullying from the respondent and also from the reports of their peers on whether the respondent had bullied them (range =0–7 peers bullied); thus a bullying tie from A to B was considered present if either A nominated B as a victim or B nominated A as a bully.
	Farrington, 2011	Single-item questionnaire	Self-reported bullying (and not peer aggression in general) was measured at age 14, in four categories of a single item: definitely no, probably no, probably yes and definitely yes.
	Hemphill, 2011	Modified version of the Communities that Care survey	Students were asked if they had taken part in "bullying another student(s) at school recently." Item responses ranged from *no* to *yes, most days* on a 4-point Likert-scale.
	Kaltiala-Heino, 2010	Questions derived from the WHO youth health study	An introduction specified bullying as follows: "We say a pupil is being bullied when another pupil, or s group of pupils, say or does nasty things to him or her. It is also bullying when a pupil is being teased repeatedly in a way she or he does not like. But it is not bullying when two pupils of about the same strength quarrel or fight.' Thereafter the respondents were asked how often they had bullied others. The response alternatives for each question were: *many times a week*, *about once a week*, *less frequently* and *not at all*. In the analyses, the responses to the bullying questions were dichotomised to "many times a week or about once a week" (= frequent involvement) versus "less frequently or not at all".
	Kendrick, 2012	Measure originally developed by Alsaker and Brunner (1999)	With reference to the current semester at school, participants were asked the following three questions: 'Have you said nasty things, mocked or teased anyone in an unpleasant way at school?'; 'Have you beaten, kicked, or assaulted anyone in an unpleasant way at school or on the way to or from school?'; and

			'Have you participated in ostracizing someone?'. Each item was rated on a four-point scale, with possible responses being: (1) 'No, it has not happened'; (2) 'Yes, it has happened once or twice'; (3) 'Yes, it has happened about once a week'; and (4) 'Yes, it has happened several times a week'. Notably, while this measure does capture many aspects of bullying, including verbal, physical, and relational bullying, it does not necessarily reflect an imbalance of power between the bully and victim, which is often included in the standard conceptualization of bullying.
	Klomek, 2008	Questionnaires filled in by a parent, a teacher and child him/herself	The children were asked about bullying other children at age 8. The alternatives were: 1 = I do not usually bully other children, 2 = I sometimes bully other children, 3 = I bully other children nearly every day. Similar questions focusing on bullying were included in parent and teacher questionnaires, with probe and response items worded as follows: The child bullies other children: 1. doesn't apply, 2. applies somewhat and 3. certainly applies. The extent of bullying at age 8 was investigated by pooling the information obtained from self-reports, parents and teacher. Combining the parent, teacher and child reports is justified by the finding that the inter-rater agreement was low (weighted kappa in range 0.11–0.22), indicating that each informant is needed to obtain a comprehensive assessment of bullying behaviour. Respondents at age 8 were classified as never bullying, bullying sometimes or bullying frequently. Sometimes bullying or was considered to exist if it was reported by at least one informant as occurring sometimes. Similarly, frequent bullying or victimization was considered to exist if it was reported by at least one informant as occurring frequently.
	Luukkonen, 2009	KSADS-PL: semi-structured Schedule for Affective Disorder and Schizophrenia for School-Age Children Present and Lifetime	Bullying others was gathered from KSADS-PL criteria for conduct disorder, and was defined as present if an adolescent had threatened or intimidated another on 3 or more occasions.
	McVie, 2014	Adapted from the Olweus Bully/Victim questionnaire	Children were asked how often they had done any of the following things to someone that they knew (not including siblings): "ignored them on purpose or left them out of things," "said nasty things, slagged them or called them names," "threatened to hurt them," or "hit, spat or threw stones at them." on a 4-point scale (3 = most days, 2 = at least once a week, 1 = less than once a week, or 0 = never). The resultant scores ranged from 0 to 12 representing a measure of frequency of bullying perpetration during the previous year for ages 13, 14, 15, and 16. The only exception to this is the bullying perpetration measure at age 13,

			which asked whether they had done each of these things (yes/no), so provides prevalence rather than a frequency measure. The "high bully at age 13–16" measure was constructed by creating a composite measure of bullying from the scales at ages 13, 14, 15 and 16 and then creating a binary variable with the highest quartile in each measure = 1.
	Pellegrini, 2001	Olweus' self-report The Senior Bully-Victim Questionnaire	Scores for bullying (six items, e.g., How often have you taken part in bullying other students in school?) were derived, following procedures recommended by Olweus. Each item had a response choice ranging from zero to four. The unit of analysis was the mean score of each item summed across the six items.
	Stallard, 2013	Single item	Participants were asked whether they had taken part in bullying others (categorised as never vs once or twice or more than 2–3 times per month) over the preceding 6 months.
	Winsper, 2012	Modified Bullying and Friendship Interview Schedule (child) + Single question (mother)	Children were asked five questions (for giving and receiving) pertaining to experience of overt bullying, specifically, whether they had experienced any of the following: had personal belongings taken; had been threatened or blackmailed; had been hit or beaten up; had been tricked in a nasty way; had been called bad/nasty names. In addition, they were asked four questions (for giving and receiving) pertaining to relational bullying: exclusion to upset the child; pressure to do things s/he didn't want to do; lies or nasty things said about others; games spoiled. Mother-reported victim status was constructed from a single question, "child is picked on or bullied by other children," asked repeatedly at 4, 7, and 9 years. Bully status was constructed from the question, "In the past year has the child bullied or threatened someone?" asked at 4, 7, and 9 years. Victim and bully status were coded as present if the mother replied "applies somewhat" or "certainly applies" at any time point. The following mother-reported bullying variable was constructed: not involved in bullying; bully/victim status; pure victim status; or pure bully status.
Aggression and/or disruptive behaviour			
Peer aggression and/or disruptive behaviour	Prinstein, 2004	Peer nominations	The girls nominated three same-sex classmates who "start fights," "interrupt," and are "bossy." A mean of standardized scores for these three items was computed as a measure of aggression and/or disruptive behaviour.

Thematic category: Victimization (e.g. bullying, aggression)

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Christina, 2021, Australia	Systematic review and meta-analysis of 73 prospective cohort studies	102189 school-aged children (youngest mean age: 6.3 years (SD not available), oldest mean age: 16 years (SD not available))	Peer victimization (mainly self-reported; multiple types, cyber victimization, physical, verbal and/or reputational victimization)	Outcomes: - Internalizing symptoms: e.g. depression, anxiety, broad internalizing symptoms (mainly using shortened or modified versions of previously validated and published measures; mainly self-reported) Statistics: Random-effects meta-analysis with meta-regression for subgroup analyses
Fisher, 2016, USA	Systematic review and meta-analysis of 55 studies (mostly cross-sectional studies, as the authors state that "the most significant limitation to this study is the use of cross-sectional data" and "Very few primary quantitative studies have examined the longitudinal impacts of peer cybervictimization, making a meta-analysis of longitudinal studies premature")	257678 adolescents aged 12-18 years attending middle or high schools (at least 50% of the sample in the USA)	Peer cyber victimization (no information available on to what extent victimization was self-, peer-, teacher- or parent-reported)	Outcomes: - Suicidal ideation - Depression - Anxiety - Self-esteem - Satisfaction - Fear - Self-harm - Aggression (no information available on to what extent these outcome measures were self-, peer-, teacher- or parent-reported) Statistics: Random-effects meta-analysis with robust variance estimation and meta-regression for subgroup analyses

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
12-18 years				
Suicidal ideation	Peer cyber victimization	Statistically significant: £ r: 0.37, 95%CI [0.23;0.51] (p<0.001) *with harm for peer cyber victimization*	Number of studies and participants unclear	Fisher, 2016
Depression		Statistically significant: £ r: 0.35, 95%CI [0.26;0.44] (p<0.001) *with harm for peer cyber victimization*		
Anxiety		Statistically significant: £		

		r: 0.31, 95%CI [0.22;0.41] (p<0.001) *with harm for peer cyber victimization*		
Self-esteem		Statistically significant: £ r: 0.21, 95%CI [0.15;0.28] (p<0.001) *with harm for peer cyber victimization*		
Satisfaction		Not statistically significant: £ r: 0.22, 95%CI [-1.04;1.49] ££† (p>0.05)		
Fear		Not statistically significant: £ r: 0.24, 95%CI [-0.08;0.55] ££† (p>0.05)		
Self-harm		Statistically significant: £ r: 0.34, 95%CI [0.20;0.48] (p<0.001) *with harm for peer cyber victimization*		
Aggression		Not statistically significant: £ r: 0.19, 95%CI [0.00;0.38] ££† (p>0.05)		
5-18 years				
All internalizing symptoms	All types of peer victimization	Statistically significant: £ r: 0.18, z: 8.31, 95%CI [0.14;0.22] (p<0.001) *with harm for peer victimization*	73, 102189	Christina, 2021
	Unspecified or general victimization	Statistically significant: £ r: 0.19, z: 7.76, 95%CI [0.14;0.24] (p<0.001) *with harm for unspecified/general victimization*	63, number of participants unclear	
	Cyber victimization	Statistically significant: £ r: 0.12, z: 7.01, 95%CI [0.08;0.15] (p<0.001) *with harm for cyber victimization*	6, number of participants unclear	
	Physical victimization	Statistically significant: £ r: 0.10, z: 2.70, 95%CI [0.03;0.16] (p=0.01) *with harm for physical victimization*	2, number of participants unclear	
	Relational victimization	Statistically significant: £ r: 0.08, z: 5.43, 95%CI [0.05;0.11] (p<0.001) *with harm for relational victimization*	2, number of participants unclear	
Anxiety	All types of peer victimization	Statistically significant: £ r: 0.31, z: 3.53, 95%CI [0.14;0.47] (p<0.001) *with harm for peer victimization*	6, number of participants unclear	
Depression		Statistically significant: £ r: 0.21, z: 8.33, 95%CI [0.16;0.25] (p<0.001) *with harm for peer victimization*	33, number of participants unclear	
General measures of internalizing symptoms		Statistically significant: £	34, number of participants unclear	

| | | r: 0.13, z: 3.67, 95%CI [0.06;0.20] (p<0.001) *with harm for peer victimization* | | |

r: Pearson's correlation coefficient
£ No raw data available
££: No SE and/or CI available; or no information on magnitude of effect available to assess variability of results.
†: Imprecision (lack of data)

Study limitations

Author, Year	Information about 'Study limitations' from the SRs
Christina, 2021	The Kmet, Cook and Lee (2004) Checklist for Assessing the Quality of Quantitative Studies was used. This contains 14 criteria including the appropriateness of the study design, sample size, methods and analyses, if random allocation and blinding of participants and investigators was carried out (where relevant), whether the aims, methods used, means, and effect sizes were reported and sufficiently described, and whether the conclusions made were supported by the results. The overall quality of the included studies was good with scores ranging from 0.64 to 1 with a mean value of 0.88 (SD =0.08) out of a possible score of zero to one. Of the 85 studies included in the meta-analysis, 84 were deemed to be of adequate quality (scores ≥ 0 .70) with one study falling just below the acceptable level. Therefore, findings from the meta-analyses conducted in the current research can be interpreted with some confidence, as the methodological quality of the primary research was generally high.
Fisher, 2016	No quality appraisal done by the systematic review authors

Certainty of the body of evidence

	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very Low [D]**	

Conclusion	**12-18 years** There is limited evidence from one systematic review with harm for peer cyber victimization: It was shown that peer cyber victimization is associated with a statistically significant increased risk of suicidal ideation, depression, anxiety, self-esteem problems and self-harm (Fisher 2016). However, a statistically significant association between cyber victimization and a change in satisfaction (across several domains), fear and aggression could not be demonstrated (Fisher 2016). Evidence is of very low certainty. **5-18 years** There is limited evidence from one systematic review with harm for peer victimization: It was shown that peer victimization (the combination of unspecified or general victimization, cyber victimization, physical victimization and relational victimization) as well as each of these individual types of peer victimization, resulted in a statistically significant increased risk of internalizing symptoms (the combination of anxiety, depression and general measures of internalizing) as well as each of these individual internalizing symptoms in children between 6 and 16 years old (Christina 2021). Evidence is of very low certainty.

	Systematic reviews
Reference(s)	Christina S, Magson NR, Kakar V, Rapee RM. *The bidirectional relationships between peer victimization and internalizing problems in school-aged children: An updated systematic review and meta-analysis.* Clin Psychol Rev 2021, 85:101979. doi: 10.1016/j.cpr.2021.101979. Fisher BW, Gardella JH, Teurbe-Tolon AR. *Peer Cybervictimization Among Adolescents and the Associated Internalizing and Externalizing Problems: A Meta-Analysis.* J Youth Adolesc 2016, 45(9):1727-43. doi: 10.1007/s10964-016-0541-z.

Measurement instruments for risk factors

Risk factor	Author, year	Instruments
Peer victimization	Christina, 2021	The most commonly used measures of peer victimization were the Social Experience Questionnaire (Grotpeter & Crick, 1996), variations of the Olweus Bully-Victim Questionnaire (Olweus, 1994), and the Peer Experiences Questionnaire (De Los Reyes & Prinstein, 2004). Almost a third (31%) of all studies used unpublished or study specific unnamed bullying measures comprising a range of one to 12 items.
Cyber victimization	Fisher, 2016	Measured either the extent to which respondents had experienced peer cybervictimization (a continuous measure) or whether or not they had experienced peer cybervictimization (a dichotomous measure). The definition of peer cybervictimization used in this meta-analysis was aggression communicated online intended to harm an individual of a similar age or social position.

Thematic category: Quality and intensity of relationships

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Averdijk, 2014, Switzerland	Observational: prospective cohort study (3 timepoints with a 1-year interval)	1675 children from 56 public primary schools, aged 7-9 years (mean age at T1 7.5±0.4 years).	- Positive relationship with siblings: assessed at T2 by asking "[child] and [sibling] play with each other" and "[child] and [sibling] get along with each other", rated by parents. - Positive relationship with friends: assessed by the Social Behavior Checklist (rated by parent, teacher and child)	Outcomes measured at T3: Internalizing behaviour (anxiety and depression) measured with the Social Behavior Questionnaire, assessed by the children, parents and teachers. Statistics: Multivariate regression
Borschmann, 2020, Australia	Observational: prospective cohort study (4 timepoints with a 1-year interval)	1239 participants, aged 8-9 years at T1, 11-12 years at T4. Mean age unknown.	Conflict with peers, assessed at T1, T2 and T3: assessed by asking "How often do you argue or fall out with your friends?"	Outcomes measured at T4: Self-harm: assessed by asking "In the past 12 months have you ever hurt yourself on purpose or done anything that might have harmed you or even killed you?" If participants responded "yes", they were then asked to describe what they did. Self-

				harm was defined as any behaviour fitting into one of five categories: (1) cutting/burning, (2) self-poisoning, (3) self-battering, (4) non-recreational risk-taking (e.g., reckless behaviour near traffic), or (5) other self-harm. Statistics: Multivariate model.
Bowes, 2010, UK	Observational: prospective cohort study (4 timepoints in a 7-year period)	1116 families with same-sex twins who were 5 years old at baseline. Follow-up home visits were conducted when children were 7 years, 10 years and 12 years. Mean age unknown.	Sibling warmth: assessed by asking mothers a series of questions about the quality of their children's relationship with one another when the children were aged 7 and 10.	Outcomes measured: Emotional and behavioural resilience to bullying victimization, derived by regressing averages scores of emotional/behavioural problems at ages 10 and 12 on levels of bullying victimization during primary school. Emotional/behavioural problems at ages 10 and 12 were assessed with the Child Behavior Checklist for mothers and the Teacher's Report Form. Statistics: Linear regression models
Branstetter, 2011, USA	Observational: prospective cohort study (2 timepoints with a 1-year interval)	166 adolescents in the 10th grade at baseline, aged 14-16 years (mean age 15.3 years, SD not available), attending public schools in both urban and suburban areas of a large metropolitan city in the Western United States.	Negative interactions with a friend at T1: 9 items of a modified version of the Network of Relationships Inventory (NRI)	Outcomes measured with Drug Involvement Scale for Adolescents (DISA; self-report): - Marijuana use - Hard drug use (cocaine, opiates, depressants, tranquilizers, hallucinogens, inhalants, stimulants, over-the-counter drugs, and club drugs) Statistics: Hierarchical regression analyses
Brendgen, 2004, Canada	Observational: prospective cohort study (2 timepoints, fall – spring during same schoolyear)	1149 participants from 37 schools in low to average SES areas in Montreal, Canada, aged 8-13 years (mean age 10.3 years) at T1. 819 participants completed questionnaires at both timepoints.	Friendship quality at T1: short version of the Friendship Quality Questionnaire (FQQ). Friendship quality residual = residual score after regressing self-perceptions onto peer-/friend-perceptions	Outcomes measured at T1 and T2: - aggression: assessed with 5 items from the Pupil Evaluation Inventory and from the Indirect Aggression Scale - depression: assessed with the

65

				Children's Depression Inventory Statistics: Hierarchical multiple linear regression analysis
Buck, 2012, USA	Observational: prospective cohort study (2 timepoints with a 3-year interval)	940 children in 6th and 9th grade, mean age 12 years at T1 and 15 at T2, from 10 sites across the USA	Friendship quality at T1 and T2: 21-item Friendship Quality Questionnaire (FQQ)	Outcomes measured at T1 and T2: - Depressive symptoms: 10-item Children's Depression Inventory Short Form Statistics: Path analysis was used to test the direct and indirect relations
Buist, 2010, The Netherlands	Observational: prospective cohort study (3 timepoints with a 1-year interval)	249 sibling pairs, aged 11-15 years (Mean age at T1 was 12.4 years for younger and 14.5 years for older siblings, SD was not available), recruited from 23 Dutch municipalities. Of these sibling pairs, 59 were brother pairs, 65 were sister pairs, 63 were older brother/younger sister pairs, and 62 were older sister/younger brother pairs.	Sibling relationship quality at T1-T3: 10-item Inventory of Parent and Peer Attachment (IPPA)	Outcomes assessed at T1-T3: - Delinquent behavior: 5-item Delinquency scale of the Nijmegen Problem Behavior List (self, sibling, father and mother) Statistics: Latent growth curve modeling
Cappadocia, 2013, Canada	Observational: prospective cohort study (2 timepoints with a 1-year interval)	1972 Canadian high school students (grade 9, 10, 11 at first timepoint) from 16 schools. No age range or mean age reported.	Quality of communication with friends: assessed by asking "How easy is it for you to talk to the following persons about things that really bother you?" with respect to their best friend, friends of the same sex, and friends of the opposite sex. Responses were given via a 5-point Likert-type scale that ranged from strongly agree to strongly disagree.	Outcomes measured at T1 and T2: - cyberbullying and cybervictimization: assessed by asking them to report how often they were involved in cyberbullying and cybervictimization in the past 2 months. Statistics: Binominal Logistic Regression
Costello, 2020, USA	Observational: prospective cohort study (2 timepoints with a 1-year interval)	184 7th (mean age 16.35±0.87 years) and 8th (mean age 17.32±0.88 years) graders from suburban and urban populations in the Southeastern United States	Intensity of observed social interactions with peers: 2 observed social interactions and coded with the Autonomy-Relatedness Coding System for Adolescent Peer Dyads and the Supportive Behavior Task Coding	Outcomes measured at T1 and T2: - depressive symptoms: assessed with the Childhood Depression Inventory - Aggressive symptoms: assessed with Youth Self Report

			System or Adolescent Peer Dyads	Statistics: Hierarchical regression models
Defoe, 2013, The Netherlands	Observational: prospective cohort study (4 timepoints with a 1-year interval)	497 adolescents (mean age 13.03±0.52 years, at baseline)	Negative interactions with a sibling or friend: Negative Interaction' subscale of the Network of Relationships Inventory (NRI)	Identified from reference list from Defoe 2013 Outcomes measured: - General mental health: externalizing problems: assessed with Youth Self Report Statistics: Cross-lagged panel analyses
Elmore, 2010, USA	Observational: prospective cohort study (2 timepoints with a 1-year interval)	587 children from 5 middle schools in 2 school districts in a southeastern US city. The resulting sample was composed of 200 3rd-grade students, 180 7th-grade students, and 185 8th-grade students. At time 2, 419 original students completed the questionnaires.	Peer attachment: the Inventory of Parent and Peer Attachment (IPPA)	Outcomes measured at T1 and T2: Aggression/resistance: assessed with the Assessment of Behavioral Disaffection Scale. Statistics: Hierarchical multiple regression analyses.
Glazebrook, 2015, UK	Observational: prospective cohort study (2 timepoints with a 6-month interval)	52 adolescents, aged 12 to 17 years (median age 15; IQR 15-16), referred to specialist child and adolescent mental health services, with a history of self-harm behavior. The most frequently endorsed methods of self-harm were self-poisoning (n = 44, 85%), self-cutting (n = 39, 77%), and battering or hitting oneself (n = 29, 57%). Three quarters of participants (n = 39, 75%) had scores of 8 or above on the anxiety subscale of the HADS, indicating probable clinical anxiety, while a third (n = 17, 33%) had scores of 8 or above on the depression subscale of the HADS, indicating probable clinical depression. A third of participants (n =	Peer attachment at T1: The Attachment Questionnaire for Children (AQC)	Outcomes measured at T2: - Self-harm: questionnaire developed from the questions used on the Hawton et al.'s (2002) large school-based survey investigating self-harm in adolescence (self-report) Statistics: Multivariable logistic regression (enter method) was used to examine those factors that independently and most strongly predicted self-harm behavior at 6-month follow-up

		17, 33%) met the criteria for probable clinical depression and anxiety.		
Hazel, 2014, USA	Observational: prospective cohort study (5 timepoints with a 3-month interval)	675 youth, recruited from metropolitan school districts in New Jersey and Colorado during the 3rd (7–9 years old), 6th (10–12 years old), and 9th (13–16 years old) grades. Mean age is not available.	Peer stressors at T1-T5: 8 items from the Adolescent Life Events Questionnaire (ALEQ)	Outcome measured at T1-T5: - Depressive symptoms: 27-item Children's Depression Inventory (CDI) Statistics: Three mixed effects models
Hedeland, 2016, Denmark	Observational: case-control study	Cases: 381 patients, 10-17 year old (mean age 14.8 years, SD unknown), admitted to 17 paediatric departments across Denmark as a result of a suicide attempt with acetaminophen. Age-matched controls: 296 healthy adolescents, recruited from 5 public schools and 2 high schools in different socio-economic areas in the Capital and Zealand regions of Denmark.	Relationship with friends and siblings: assessed with self-developed scale, no clear information on questions asked	Outcomes measured: suicide attempts Statistics: Fisher exact test
Kidd, 2006, Canada	Observational: prospective cohort study (2 timepoints with a 1-year interval) (first 2 timepoints of Add Health)	9142 adolescents in grades 7 through 11, median age 16 years (mean age or age range unknown)	Peer relations at T1 and T2: adolescents reported on 5 items assessing the frequency with which they have spent time, talked and discussed their problems with their closest male and female friends.	Outcomes measured at T1 and T2: - Suicide attempts (adolescents were asked how many times they had attempted suicide in the past year) Statistics: Hierarchical logistic regression
Kingery, 2007, USA	Observational: prospective cohort study (2 timepoints in a 6-month period)	146 children, aged 10-11 (grade 5; mean age at T1 11 years and 1 month). Elementary and middle schools from two public school districts located in low- to middle-income communities in northern New England were recruited.	Friendship quality at T1: 40-item Friendship Quality Questionnaire-Revised (FQQ-R)	Identified from reference list from Grills-Taquechel 2010 (see evidence summaries "Social support" & "Being loved and being part of the group") Outcomes measured at T2: - Loneliness: Loneliness and Social Dissatisfaction Questionnaire, self-reported - School involvement: Attitudes Toward

| | | | | School self-reported questionnaire

Statistics: Hierarchical regression analyses |
|---|---|---|---|---|
| Kingery, 2011, USA | Observational: prospective cohort study (2 timepoints in a 6-month period) | 365 adolescents, aged 10-11 (grade 5; mean age at T1 11 years and 2 months). Elementary and middle schools from six public school districts located in lower- to middle-income rural and suburban communities in northern New England were recruited. | Friendship quality at T1: 40-item Friendship Quality Questionnaire-Revised (FQQ-R) | Identified from reference list from Vanhalst 2013 (see evidence summary "Being loved and being part of the group")

Outcomes measured at T2:
- Loneliness: Loneliness and Social Dissatisfaction Questionnaire, self-reported
- Depression: The Children's Depression Inventory, self-reported
- School involvement: Attitudes Toward School self-reported questionnaire
- Self-concept: Self-Perception Profile for Children. Only the general self-worth domain was used in the present study.

Statistics: Regression analyses |
| Koch, 2020, USA | Observational: prospective cohort study (3 timepoints with a 4-month interval)
The 4-month intervals were selected to collect responses throughout the course of a school year. Accordingly, baseline corresponded with the summer, T2 corresponded with the fall and T3 with the spring of the | 188 girls, aged 9-14 years at baseline (mean age 11.70±1.05 years) recruited through advertisement via emails and canvasing parents at drop-off for summer youth activity programs. Participants attended several schools in the Upstate New York area and represented a cross-section of middle school grades. | Peer problems at T1: 25-item Index of Peer Relations (IPR) | Outcomes measured at T3:
- Depressive symptoms: 20-item self-report Center for Epidemiological Studies Depression Scale for Children (CES-DC)

Statistics: Path analysis (structural equation modeling) |

		same school year.			
Laird, 2013, USA	Observational: prospective cohort study (part of a larger 3-timepoint longitudinal study) [only second (T1) and third (T2) annual timepoints used in this study]	181 adolescents from public schools in Southern USA, mean age 12.4±0.53 years at T1; 148 adolescents at T2 (mean age 13.4±0.56 years)	Friendship quality: selected items from the Friendship Quality Scale (FQS).	Outcomes measured: - depression during the past month: 6-item Modified Depression Scale (self-report) - antisocial behaviour during the past month: 26 items from the Problem Behavior Frequency Scale (self-report) Statistics: Polynomial regression equations
Le Grange, 2014, Australia	Observational: prospective cohort study (2 timepoints with 2-3 year interval, part of larger study with 15 timepoints)	1300 youth, aged 12-16 years (mean age was not available), recruited through selected Maternal and Child Health Centers across both urban and rural areas in the State of Victoria, Australia	Peer relationships at 13-14 years: Inventory of Parent and Peer Attachment (IPPA)	Identified from reference list from Pace 2021 Outcomes measured: - Abnormal Eating Attitudes and Behaviors at 15–16 years of age based on the Drive for Thinness and Bulimia Subscales of the Eating Disorder Inventory (EDI), and an adapted version of the EDI Body Dissatisfaction Subscale suitable for administration in a (non-clinical) general population sample Statistics: Path analysis
McNeil, 2020, USA	Observational: prospective cohort study (3 timepoints with 1.5 year interval)	657 adolescents, aged 11-14 year (mean age 12.49±1.13 at baseline), were subsampled from National Survey of	Peer relationships at T1-T3 (rate of change): 16-item Loneliness and Social Dissatisfaction	Outcomes measured T1-T3 (rate of change): - Depressive symptoms: 27-item Children's Depression inventory (CDI)

		Child and Adolescent Well-Being (NSCAW-II).	Questionnaire for Young Children	Statistics: Latent growth models
Meter, 2015, USA	Observational: prospective cohort study (2 timepoints with a 1-year interval)	243 students from a middle school in a southwestern US city, aged 11-14 years (mean age 12.3 years, SD unknown)	Friendship qualities at T1: after choosing a best friend, the participants then rated how often certain things happened within their relationship with the chosen friend, in order to describe the construct of the quality of intimate exchange ("We tell each other our problems"; "Talk about things that make us upset"; "Tell each other secrets"). A construct of perception of negative friendship quality, referred to as conflict ("We argue;" We get mad at each other;" "We disagree about things"), was included in the analyses to control for the effect this negative friendship characteristic may have on peer influence effects within a friendship.	Outcomes measured at T1 and T2: Peer-nominated aggression: The construct included proportion scores from questions that asked the participants to identify their peers who hit or kick others, push others around, and say mean things to others or call them names (3 overt aggression items), and also those who try to keep others from being part of activities or groups, spread rumours about others, or try to make others ignore or not talk to certain peers (3 relational aggression items). Statistics: Hierarchical regression analysis
Moser, 2002, USA	Observational: prospective cohort study (2 timepoints with a 3-year interval)	99 intact families were recruited from the San Francisco Bay area, all of whom contained both parents and at least one target child between 10 and 18 years of age (mean age 13.8±1.6 years) still living in the home. The siblings were on average 12.8±2.3 years.	Sibling warmth & conflict at T1: Sibling Relationship Questionnaire	Identified from reference list from Branje 2004 (see evidence summary "social support") Outcomes measured at T2: - Internalizing and externalizing behaviour: Child Behaviour Checklist Statistics: Autoregressive prediction model
Munroe, 2020, USA	Observational: prospective cohort study (2 timepoints, the average period between T1 and T2 data collection was 2.84 years (M= 1038.24 days;	45 primary caregivers of children who were transgender and gender-diverse (TGD, including children with binary transgender, non-binary transgender, and	Peer relations: the peer relations subscale of the Child Behaviour Checklist, parent version (CBCL). This subscale was created from items 25 ("Does not get along with other kids"), 38 ("Gets teased a lot"), and 48	Outcomes measured at T1 and T2: - Internalizing and externalizing behaviour: Child Behaviour Checklist, parent version (CBCL) Statistics: Longitudinal path model

71

	SD = 77.81; Median = 1063.78 days)	gender-diverse identities and expressions). Children ranged in age from 6-12 years (mean age 8.5±1.5 years).	("Not liked by other kids").	
Nelemans, 2016, The Netherlands/ USA	Observational: prospective cohort study (6 timepoints with a 1-year interval)	497 adolescents, mean age 13.03±0.46 years at baseline, attending the first grade of secondary schools in western and central regions of the Netherlands at the start of the study	Peer involvement at T1-T6: 5-item "intensity of contact with friends" subscale of the Questionnaire on Peer Relationships	Outcomes measured at T1-T6: - Addiction: cannabis use: "In the past 12 months, how often have you used weed, marihuana or hashish?". Adolescents rated the item on a 14-point scale, ranging from 0 (zero times) to 10 (10 times), followed by 11 to 19 times, 20 to 39 times, and 40 times or more. - Social anxiety disorder (SAD) symptoms: 4-item SAD subscale of the Dutch version of the original 38-item Screen for Child Anxiety Related Emotional Disorders (SCARED) Statistics: Longitudinal cross-lagged panel model
Nelemans, 2017, The Netherlands	Observational: prospective cohort study (7 timepoints with a 1-year interval)	631 youths, mean age 7.96±0.35 years at baseline, recruited in second grade to participate in the Social Health and Relationship Experiences project, a longitudinal study of peer relationships in the Midwest area of the United States	Friendship quality at T5-T7: 25 items of the Inventory of Parent and Peer Attachment (IPPA)	Identified from reference list from Barzeva 2002 (see evidence summary "Being loved and being part of the group") Outcomes measured at T1-T7: - Anxiety symptoms: 28-item Revised Children's Manifest Anxiety Scale (self-report) Statistics: Growth mixture modeling
Nrugham, 2008, Norway	Observational: prospective cohort study (2 timepoints with a 1-year interval)	2464 adolescents with mean age 13.7±0.5 years (age range 13-14 years) at T1. At T2, 2432 adolescents, completed the same questionnaire again.	Peer attachment at T1 and T2: Inventory of Parent and Peer Attachment (IPPA)	Outcomes measured at T1 and T2: Suicidal acts, assessed via questionnaire and interviews: "Have you ever tried to commit suicide?" ("No, never"; "Yes, once";

				"Yes, several times"). If positively endorsed, the participants were further asked to fill out details of the timing of the last act: 'How long ago was your last act of attempted suicide?' ("Years", "Months"). Questions on suicidal behaviour were also parts of the screening probes for depression in the K-SADS-PL. Statistics: All variables with significant Odds Ratios (ORs) in the bivariate logistic regression analyses were entered into a subsequent multivariate logistic regression analysis.
Pace, 2021, Italy	Observational: case-control study	56 female adolescents, aged 14 – 18 years (mean age 16.4±1.3 years) were considered part of the Binge eating (BE) group and matched for age and gender with 56 peers without BE (mean age 16.4±1.3 years).	- Peer relations, intended as quality of best friendship - Sibling relations, in terms of warmth, hostility and rivalry Assessed with the Friends and Family Interview (FFI)	Outcome measured: Binge eating, assessed with the Italian version of the Binge Eating Scale (BES) Statistics: Multiple regression analysis
Patalay, 2018, UK	Observational: prospective cohort study (2 timepoints with a 3-year interval)	9553 participants, aged 11-14 years, are from the Millennium Cohort Study, a UK wide birth cohort study of just over 19,500 individuals born at the start of the millennium (Sept. 2000–Jan. 2002)	Peer problems: subscale of Strengths and Difficulties Questionnaire (SDQ) Argues with friends: asking participants how often they argued or fell out with their friends (1 = at least once a week/have no friends)	Outcomes measured: - Depressive symptoms at T2 (self-report): 13-item Short Moods and Feelings Questionnaire - Mental wellbeing at T1 & T2 (self-report): 6-item measure assessing satisfaction with different aspects domains which are appropriate for children including school, family, friends, school work, appearance, and life as a whole. Participants responded by indicating their level of happiness with each aspect of their lives. - Emotional symptoms at T1 & T2 (parent-report): the emotional

				symptoms and conduct problems subscales of the Strengths and Difficulties Questionnaire (SDQ) Statistics: Regression models
Pisarska, 2020, Poland	Observational: prospective cohort study (2 timepoints with a 2-year interval)	551 adolescents in the 10th and 12th grade, aged 16-18 years, attending public/non-public general, technical high schools and basic vocational schools from Warsaw	Peer relationships at T1 (including quality of relationship with peers, assertiveness and communication). Little if any further information is readily available on the scale used (see Okulicz-Kozaryn, 2001, Polisch article).	Outcomes measured: - gambling involvement that includes both gambling-related behaviours and problems (self-report) Statistics: Generalised linear models with gamma variation
Pittenger, 2018, USA	Observational: prospective cohort study (3 timepoints with a 18-month interval)	763 adolescents, aged 11 to 17 years, involved in child welfare services	Peer relationships at T1: 16-item Loneliness and Social Dissatisfaction Questionnaire (LSDQ)	Outcomes measured at T1 and T3 with Youth Risk Behavior Survey (YRBS) - past 30-day binge alcohol (on how many days did you have 5 or more drinks of alcohol in a row, that is, within a couple of hours?) - past 30-day marijuana use - past 30-day cocaine use Statistics: The Weighted Least Squares with Mean and Variance estimator (WLSMV) using full information maximum likelihood was used to run three logistic regression models predicting the outcomes of interest.
Power, 2005, USA	Observational: prospective cohort study (4 timepoints with a 6-month interval)	1253 adolescents in the 9th to 12th grade from six high schools in a large metropolitan school district in the Southwest	Peer security: Inventory of Parent and Peer Attachment	Outcomes measured: - Problematic drinking: Quantity and frequency of consumption were tapped by questions on the frequency of drinking in the past year and month, frequency of intoxication, typical quantity of consumption and frequency of consumption of large quantities.

				Statistics: Logistic regression models
Prinstein, 2005, USA	Observational: prospective cohort study (3 timepoints with a 1-year interval)	520 children and adolescents in grades 6-8, mean age boys 12.68±0.98 and girls 12.63±0.91 at baseline, enrolled in public schooling within a city of fairly homogeneous middle-class socioeconomic status	Friendship quality at T1 & T2: 7 subscales from the Network of Relationships Inventory (NRI)	Outcomes measured at T3: - Depressive symptoms: 27-item The Children's Depression Inventory (self-report) Statistics: Hierarchical regression analysis
Selfhout, 2009, The Netherlands	Observational: prospective cohort study (2 timepoints with a 1-year interval)	307 Dutch middle adolescents, aged 14 to 17 years (mean age 15.5±0.6 years) at baseline, attending 12 high schools in the province of Utrecht.	Perceived friendship quality: Support scale of the Network of Relationship Inventory (NRI)	Outcomes measured at T1 & T2: - Depressive symptoms: 27-item The Children's Depression Inventory (self-report) - Social anxiety: Social Anxiety subscale of the revised version of the Screen for Child Anxiety Related Emotional Disorders (SCARED) Statistics: Hierarchical regression analyses
Sund, 2002, Norway	Observational: prospective cohort study (2 timepoints with a 1-year interval)	2465 adolescents, aged 12.5-15.7 years (mean age 13.7±0.58 years), attending 22 schools in two counties in the middle of Norway, South and North Trøndelag	Attachment to peers: 25-item revised version of Inventory of Parent and Peer Attachment (IPPA)	Outcomes measured at T1 & T2: - Severe depressive symptoms: 34-item Mood and Feelings Questionnaire (MFQ). Total scores range between 0 and 68 and the MFQ variables were dichotomized at MFQ = 33. Statistics: logistic regression analysis
Tu, 2020, USA	Observational: prospective cohort study (2 timepoints with a 7-month interval)	100 adolescents, mean age 11.05±0.33 at baseline, in the 5th grade participated. Their mothers and their 5th grade teachers (78 teacher reports) participated as well. Families were recruited from the Mid-western United States across two consecutive cohorts.	Positive peer relationships: - 25-item Friendship Quality Questionnaire (completed by adolescents) -7-9 item questionnaire on adolescents' friends' prosociality - 6-item Checklist for Peer Relations (completed by mothers and teachers)	Outcomes measured: - 10 item Sleek/wake problems subscale of the School Sleep Habits Survey (completed by adolescents) Statistics: Cross-lagged panel models

Van Zalk, 2015, Sweden	Observational: prospective cohort study (3 timepoints with a 1-year interval)	2194 adolescents (mean age 13.58 years, SD not available) recruited from a large community-based, cohort sequential study in a community with a total population of about 26,000 in Western Europe	Friend care: Validation and Caring subscale from the Friendship Quality Questionnaire (FQQ)	Outcomes measured: - Social anxiety: 8 questions about usual fears in different situations Statistics: Simulation Investigation for Empirical Network Analyses (SIENA). Effects on the outcome variable (i.e., changes in social anxiety) obtained through the SIENA estimation can be interpreted as effects in longitudinal multinomial logistic regression
Van Zalk, 2017, Sweden	Observational: prospective cohort study (3 timepoints with a 8-month interval)	526 adolescents (423 initial participants and 103 of their online friends), aged 13-15 years (mean age 14.05 years, SD unknown), attending school in a medium-sized town in Sweden	Friendship quality with best online friend: 6 questions about perceived support and trust from the Friendship Quality Questionnaire (FQQ)	Outcomes measured: - depressive symptoms: shortened version of the Child Depression Scale from the Center for Epidemiological Studies (CESD-10) - social anxiety: Social Phobia Screening Questionnaire for Children Statistics: Autoregressive cross-lagged model
Vaughan, 2010, USA	Observational: prospective cohort study (5 timepoints with 6-month intervals)	3444 students in the 6th - 8th grade, aged 12-16 years for all data points (mean age at T1 was 13.36±0.80), attending all of the 19 schools with the targeted grade levels in 3 rural counties in a South-eastern state	Friend closeness: scale to assess how close they felt to each friend	Identified from reference list from Auerbach 2011 (see evidence summary "social support") Outcomes measured: - Depressive symptoms: 13-item Short Mood and Feelings Questionnaire (SMFQ) Statistics: Growth curve modeling
Weerman, 2018, the Netherlands	Observational: prospective cohort study (6 timepoints with 2-week intervals)	178 9th grade students, aged 14-16 years (mean age 14.37±0.56 years) at baseline, from one mid-sized high school in Kentucky. Analysis based on 155 students who participated in first 5 timepoints	Dynamic peer socialization (average similarity): 'average similarity' effect in the SIENA model. This parameter estimates the extent to which respondents adjust their offending towards the average level of offending of their nominated peers in the network (i.e. the extent to which	Outcomes measured at all timepoints: Offending, assessed as minor violence, serious violence and property offending/vandalism. Students were asked to identify the number of days over the previous 2 weeks that they had participated in 10 different delinquent activities.

		they become more similar to the average of their alters)		Statistics: The SIENA (Simulation Investigation for Empirical Network Analysis) method
Wood, 2017, Canada	Observational: prospective cohort study (2 timepoints with a 4-month interval)	430 participants from grades 5 and 6, aged 10-13 years (mean age 10.87±0.73 years) from 3 mixed-sex public schools in inner suburbs of Montréal, Canada, attended by students from across a broad middle-class section of the socio-economic spectrum.	Friendship quality at T1 and T2: Intimacy subscale of the Network Relationship Inventory (NRI)	Outcomes measured at T1 and T2: Self-assessed anxiety: the participants rated 3 items intended to measure feelings of general anxiety. They were "I am nervous or tense," "I worry a lot," and "I get stressed a lot." Statistics: Structural equation modelling
Zhang, 2018, The Netherlands	Observational: prospective cohort study (4 timepoints with 6-month interval)	1126 secondary school Dutch adolescents (age range 11-18 years, mean age 13.95±1.18 at T1)	Friendship Quality: the satisfaction and conflict subscales of the Network of Relationship Inventory (NRI)	Outcome measured: Depressive mood, measured with 6 items from the Depressive Mood List Statistics: Longitudinal multilevel analysis

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
5-11 years				
Relationship with friends				
Friendship quality				
Aggression at T2	Friend-rated friendship quality T1	Not statistically significant: B: -0.10 £† (p>0.05)	1, 134 §	Brendgen, 2004
	Overestimation of friendship quality (residual) T1	Statistically significant: B: 0.12 £ (p<0.05) *With harm for overestimation of friendship quality*		
Depression at T2	Friend-rated friendship quality T1	Not statistically significant: B: -0.05 £† (p>0.05)		
	Overestimation of friendship quality (residual) T1	Not statistically significant: B: -0.11 £† (p>0.05)		
	Friendship quality T1	Not statistically significant: B: -0.56±0.48 £† (p>0.05)	1, 339 §	Kingery, 2011
Loneliness at T2		Statistically significant: B: -0.08±0.04 (p<0.05) *With benefit for friendship quality*		
		Not statistically significant: B: -0.11±0.08 £† (p>0.05)	1, 141 §	Kingery, 2007
School involvement at T2		Not statistically significant: B: 0.11±0.07 £†		

		(p>0.05)		
		Not statistically significant: B: 0.07±0.05 £† (p>0.05)	1, 339 §	Kingery, 2011
Self-esteem at T2		Statistically significant: B: 0.10±0.04 (p<0.05) *With benefit for friendship quality*	1, 331 §	

Friendship intimacy and positive relationships

Anxiety	Friendship intimacy	Not statistically significant: β: 0.05 £† (p>0.2)	1, 430	Wood, 2017
Child-reported internalizing behaviour	Positive relationship with friends (prosocial behaviour of reciprocal friend)	Not statistically significant: B: -0.049 £† (p>0.05)	1, 361 §	Averdijk, 2014
Parent-reported internalizing behaviour		Not statistically significant: B: 0.059 £† (p>0.05)	1, 360 §	

Relationships with peers

Sleep/wake problems at T2	Positive peer relationships T1	Statistically significant: B: -0.23 £ (p<0.01) *With benefit for positive peer relationships*	1, 100 §	Tu, 2020
Internalizing symptoms T2	Peer problems T1	Not statistically significant: B: 5.18±11.60 £† (p>0.05)	1, 49 §	Munroe, 2020
Externalizing symptoms T2		Not statistically significant: B: 1.81±8.28 £† (p>0.05)		
Self-harm T4	Conflict with peers (during one earlier timepoint vs no timepoint)	Not statistically significant: aOR: 2.59, 95%CI [0.59;7.06] ¥ (p>0.05)	1, 1067	Borschmann, 2020
	Conflict with peers (during 2 or 3 earlier timepoints vs no timepoint)	Not statistically significant: aOR: 2.77, 95%CI [0.89;8.66] ¥ (p>0.05)		

Relationships with siblings

Child-reported internalizing behaviour	Positive relationship with siblings	Not statistically significant: B: 0.002 £† (p>0.05)	1, 361 §	Averdijk, 2014
Parent-reported internalizing behaviour		Not statistically significant: B: -0.050 £† (p>0.05)		
Emotional resilience	Sibling warmth	Statistically significant: β: 0.15 £ (p<0.001) *With benefit for sibling warmth*	1, 1116	Bowes, 2010
Behavioral resilience		Statistically significant: β: 0.24 £ (p<0.001) *With benefit for sibling warmth*		

12-18 years

Relationships with friends

Friendship quality

Depressive symptoms at T2	Friendship quality T1	Not statistically significant: ££† (p>0.05) (model 3)	1, 940	Buck, 2012
		Not statistically significant: B: -0.02 £† (p>0.05)	1, 307 §	Selfhout, 2009
		Not statistically significant:	1, 148 §	Laird, 2013

		B: 0.01±0.09 £† (p=0.89)		
		Statistically significant: β: 0.16 £ (p<0.01) *With harm for friendship quality*	1, 526	Van Zalk, 2017
Depressive symptoms at T3	Friendship quality T2	Not statistically significant: β: -0.09 £† (p>0.05)		
	Positive friendship quality T2	Not statistically significant: B: 0.02±0.02 £† (p>0.05)	1, 438	Prinstein, 2005
Anxiety symptoms (increasing vs decreasing) at T7 (mean±SE)	Friendship quality	Statistically significant: 3.49±0.15 vs. 3.96±0.07 $ MD: -0.47 £ (p<0.036) *With benefit for friendship quality*	1, 631	Nelemans, 2017
Anxiety symptoms (high vs low) at T7 (mean±SE)		Statistically significant: 3.56±0.10 vs. 3.97±0.03 $ MD: -0.41 £ (p<0.036) *With benefit for friendship quality*		
Social anxiety at T2	Friendship quality T1	Not statistically significant: B: -0.05 £† (p>0.05)	1, 307 §	Selfhout, 2009
		Not statistically significant: β: 0.04 £† (p>0.05)	1, 526	Van Zalk, 2017
Social anxiety at T3	Friendship quality T2	Statistically significant: β: -0.15 £ (p<0.05) *With benefit for friendship quality*		
Antisocial behaviour at T2	Friendship quality T1	Statistically significant: B: 0.08±0.03 (p=0.008) *With harm for friendship quality*	1, 148 §	Laird, 2013
Cyberbullying (vs no involvement)	Lower quality of friendships	Not statistically significant: ££† (univariate result not significant, not included in multivariate model) (p>0.05)	1, 62 vs 1567	Cappadocia, 2013
Cybervictimization (vs no involvement)		Not statistically significant: B: -0.11±0.19 £† (p>0.05)	1, 95 vs 1599	
Cyberbullying & cybervictimization (vs no involvement)		Not statistically significant: B: -0.42±0.26 £† (p>0.05)	1, 54 vs 1544	
Binge eating (vs not)	Quality of contact with best friend	Not statistically significant: 3.05±1.09 vs 3.63±4.32 $ MD: -0.58 £† (p=0.929)	1, 56 cases vs 56 controls §	Pace, 2021
Friendship care, closeness and satisfaction				
Depressive symptoms	Friendship closeness (within-person effect)	Statistically significant: B: -0.07±0.03 (p<0.05) *With benefit for friendship closeness (within-person effect)*	1, 3444	Vaughan, 2010
	Friendship closeness (between-person effect)	Not statistically significant: B: -0.07±0.05 £† (p>0.05)		
Depressive mood	Friend satisfaction	Statistically significant: B: -0.15±0.03 (p<0.001)	1, 1126	Zhang, 2018

		With benefit for higher friend satisfaction		
Suicidal acts (vs no suicidal acts)	Friend attachment	Not statistically significant: OR: 1.04, 95%CI [0.97;1.1] $ (p>0.05) (not significant and therefore not included in multivariate model)	1, 37 vs 228 §	Nrugham, 2008
Changes in social anxiety	Friend care	Statistically significant: B: -0.04 £ (p<0.01) *With benefit for friend care*	1, 2194	Van Zalk, 2015
Friendship intensity				
Depressive symptoms at T2	Friendship intensity T1 (moderated by autonomous relatedness with best friend)	Statistically significant: β: 0.17, 95%CI [0.05;0.30] (p<0.001) *With harm for friendship intensity (moderated by autonomous relatedness with best friend)*	1, 184 §	Costello, 2020
Aggressive symptoms at T2		Not statistically significant: β: 0.07, 95%CI [-0.06;0.22] £† (p>0.05)		
Depressive symptoms at T2	Friendship intensity T1 (moderated by close friend engagement)	Not statistically significant: β: -0.05, 95%CI [-0.18;0.07] £† (p>0.05)		
Aggressive symptoms at T2		Not statistically significant: β: 0.016, 95%CI [-0.14;0.14] £† (p>0.05)		
Aggression at T2	Intimate exchange T1	Not statistically significant: B: 0.00±0.00 £† (p>0.05)	1, 243 §	Meter, 2015
Negative interactions with friends				
Depressive symptoms at T2	Argues with friends T1	Not statistically significant: B: -0.02±0.04 £† (p>0.05)	1, 9553	Patalay, 2018
Emotional symptoms at T2		Not statistically significant: B: 0.00±0.03 £† (p>0.05)		
Wellbeing at T2		Not statistically significant: B: -0.04±0.04 £† (p>0.05)		
Depressive mood	Friend conflict	Statistically significant: B: 0.14±0.03 (p<0.001) *With harm for more friend conflict*	1, 1126	Zhang, 2018
Marijuana use at T2	Negative interactions with a friend T1	Not statistically significant: β: 0.03 £† (p>0.05)	1, 166 §	Branstetter, 2011
Hard drug use at T2		Not statistically significant: β: -0.08 £† (p>0.05)		
Externalizing problems at T2-T4	Negative interactions with a friend	Not statistically significant: ££† (p>0.05) (combined model)	1, 497	Defoe, 2013
Relationships with peers				
Peer involvement and relationships				
Severe depressive symptoms at T2	Attachment to peers at T1	Not statistically significant: aOR: 0.99, 95%CI [0.85;1.16] (p>0.05)	1, 1973	Sund, 2002
Resistance/ aggression behaviour at T2	Peer attachment T1 (moderated by school satisfaction)	Not statistically significant: β: -0.10 £† (p>0.005)	1, 395 §	Elmore, 2010
Social anxiety disorder symptoms 1 year later	Peer involvement	Statistically significant: B: -0.07, 95%CI [-0.119;-0.025] (p=0.003)	1, 497	Nelemans, 2016

Cannabis non-use 1 year later		*With benefit for peer involvement* Statistically significant: aOR: 0.22, 95%CI [0.144;0.326] (p<0.001) *With harm for peer involvement*		
Cannabis use frequency		Statistically significant: B: 0.23 £ (p=0.04) *With harm for peer involvement*		
Suicide attempts at T2	Peer relations T1	Not statistically significant: ££† (p>0.05)	1, 9142	Kidd, 2006
Marijuana use at T3	Peer relationships T1	Not statistically significant: aOR: 0.87, 95%CI [0.69;1.10] ¥ (p>0.05)	1, 763	Pittenger, 2018
Cocaine use at T3		Not statistically significant: aOR: 0.99, 95%CI [0.68;1.45] ¥ (p>0.05)		
Binge drinking at T3		Not statistically significant: aOR: 0.86, 95%CI [0.67;1.10] ¥ (p>0.05)		
Abnormal eating attitudes & behaviors at T2		Statistically significant: β: 0.20 £ (p<0.001) *With harm for stronger peer relationships*	1, 1300	Le Grange, 2014
Gambling involvement at T2	Positive relationship with peers T1	Not statistically significant: r: -0.35 £† (univariate result not significant, not included in multivariate model) (p>0.05)	1, 261 §	Pisarska, 2020
(changes in) problematic drinking	Peer security T1	Not statistically significant: £† (univariate result not significant, not included in multivariate model) (p>0.05)	1, 743	Power, 2005
Offending	Dynamic peer socialization (average similarity)	Not statistically significant: B: -0.565±1.535 £† (p>0.05)	1, 155 §	Weerman, 2018
Problems with peer relationships or attachment				
Depressive symptoms at T2	Peer problems T1	Statistically significant: B: 0.06±0.02 (p<0.001) *With harm for peer problems*	1, 9553	Patalay, 2018
Emotional symptoms at T2		Statistically significant: B: 0.08±0.02 (p<0.001) *With harm for peer problems*		
Wellbeing at T2		Statistically significant: B: -0.03±0.01 (p<0.05) *With harm for peer problems*		
Depressive symptoms (rate of change)	Peer relationship problems (rate of change)	Statistically significant: β: 0.552±0.091 (p<0.001) *With harm for increases in peer relationship problems*	1, 657	McNeil, 2020
Self-harm at T2	Insecure/poor peer attachment T1	Statistically significant: aOR: 8.01, 95%CI [1.00;64.20] (p<0.05) *With harm for poor peer attachment*	1, 49 §	Glazebrook, 2015
Relationships with siblings				
Internalizing behaviour at T2	Sibling warmth T1	Not statistically significant: β: -0.02 £†	1, 99 §	Moser, 2002

		(p>0.05)		
Externalizing behaviour at T2		Not statistically significant: β: -0.07 £† (p>0.05)		
Binge eating (vs not)	Sibling warmth	Not statistically significant: 2.83±1.09 vs 2.97±0.66 $ MD: -0.14 £† (p=0.372)	1, 56 cases vs 56 controls §	Pace, 2021
Delinquent behavior at T3	Sibling relationship quality	Not statistically significant: ££† (p>0.05)	1, 249 §	Buist, 2010
Internalizing behaviour at T2	Sibling conflict T1	Statistically significant: β: 0.20 £ (p<0.05) *With harm for sibling conflict*	1, 99 §	Moser, 2002
Externalizing behaviour at T2		Not statistically significant: β: 0.08 £† (p>0.05)		
Binge eating (vs not)	Sibling hostility	Not statistically significant: 2.97±0.66 vs 1.84±1.00 $ MD: 1.13 £† (p=0.299)	1, 56 cases vs 56 controls §	Pace, 2021
Externalizing problems at T2	Sibling negative Interactions T1	Not statistically significant: β: -0.01±0.01 £† (p>0.05)	1, 497	Defoe, 2013
Externalizing problems at T3	Sibling negative Interactions T2	Not statistically significant: β: -0.01±0.01 £† (p>0.05)		
Externalizing problems at T4	Sibling negative Interactions T3	Not statistically significant: β: -0.01±0.01 £† (p>0.05)		
Binge eating (vs not)	Sibling rivalry	Not statistically significant: 1.84±1.00 vs 1.64±0.81 $ MD: 1.03 £† (p=0.588)	1, 56 cases vs 56 controls §	Pace, 2021
5-18 years				
Depressive symptoms	Peer stressors	Statistically significant: B: 0.147±0.035 (p<0.01) *With harm for peer stressors*	1, 675	Hazel, 2014
Depressive symptoms at T3	Peer problems T1	Not statistically significant: B: 0.11±0.17, 95%CI [-0.22;0.44] £† (p>0.05)	1, 188 §	Koch, 2020
Suicide (vs control)	Close relationship with friends	Statistically significant: 188/350 vs 240/291 $ OR: 0.25, 95%CI [0.17;0.36] (p<0.00001) *With benefit for close relationship with friend*	1, 350 cases vs 291 controls	Hedeland, 2016
	Close relationship with siblings	Statistically significant: 114/325 vs 159/278 §$ OR: 0.40, 95%CI [0.29;0.56] (p<0.00001) *With benefit for close relationship with siblings*	1, 325 cases vs 278 controls	
	Intermediate relationship with friends	Statistically significant: 99/350 vs 49/291 §$ OR: 1.95, 95%CI [1.33;2.86] (p=0.0007) *With harm for intermediate relationship with friend*	1, 350 cases vs 291 controls	
	Intermediate relationship with siblings	Not statistically significant: 96/325 vs 102/278 §$ OR: 0.72, 95%CI [0.51;1.02] ¥	1, 325 cases vs 278 controls	

	(p=0.06)			
	Dissociated relationship with friends	Statistically significant: 53/350 vs 2/291 §$ OR: 25.79, 95%CI [6.23;106.79] (p<0.00001) *With harm for dissociated relationship with friend*	1, 350 cases vs 291 controls	
	Dissociated relationship with siblings	Statistically significant: 108/325 vs 17/278 §$ OR: 7.64, 95%CI [4.44;13.14] (p<0.00001) *With harm for dissociated relationship with siblings*	1, 325 cases vs 278 controls	

Mean ± SD (unless otherwise indicated), B/β ± SE, MD: mean difference, aOR: adjusted odds ratio, SD: standard deviation, SE: standard error, B: unstandardized coefficient, β: standardized coefficient, r: correlation coefficient

$ The outcome measures and effect measures represent the risk factor/exposure, not the outcome

£ No SE and/or CI available; or no information on magnitude of effect available to assess variability of results

££ No raw data or effect size reported

¥ Imprecision (large variability of results)

† Imprecision (lack of data)

§ Imprecision (limited sample size or low number of events)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Averdijk, 2014	No, cluster randomized sampling approach was used.	No, outcome was both child- and parent-reported. Exposure was either parent-reported (relationship with siblings) or child-, parent- and teacher-reported	No, adjusted for gender, socio-economic status, migrant ethnicity status, parental conflict and aggression.	Unclear, 95% (children)/97% (parents) retention rate at T2, 95% (children)/96% (parents) retention rate at T3, no attrition analysis performed.	N/A
Borschmann, 2020	No, participants came from a stratified random sample of 43 primary schools.	Yes, outcome and exposure measures were self-reported.	No, adjusted for age, sex and SEIFA (Socio-Economic Index For Areas) quintile.	Unclear, no information on differences between those who did all 4 timepoints, and those who did not.	Yes, the sample was skewed towards higher SES and had a higher proportion of Indigenous children than the general Australian population. They did not ask about self-harm prior to T4 and this may represent a missed opportunity

83

					to collect informative data.
Bowes, 2010	No, high-risk stratification strategy used to replace any families lost to the original register at the time of birth owing to selective non-response, and a further high-risk oversample was used to ensure sufficient numbers of children growing up in adverse environments.	No, outcome and exposure were mother- and/or teacher-reported	No, adjusted for IQ, socio-economic status, gender and baseline problems.	Unclear, no information on differences between families who participated at all timepoints and those who dropped out earlier.	Yes, the sample comprised twins and we cannot be certain that the results generalise to singletons.
Branstetter, 2011	Unclear, no information on recruitment procedure (probability or non-probability sampling?)	Yes, outcome measures were self-reported. Adolescents may underreport their substance usage. However, multiple reporters (friends and mothers) to assess qualities of relationships were used.	No, hierarchical regression analyses were conducted in order to determine how friendship factors and mother-adolescent relationship quality operate together in predicting the frequency of substance use. To control for prior use, substance use frequency in the 10th grade was entered as the first step (baseline result).	No, the subsample of 166 adolescents was compared to the 34 excluded adolescents on all variables used in this study, and no differences were found.	Yes, the present study focused on a single close friendship. It would also be important to incorporate other key relationships that emerge in adolescence, including other-sex friends and romantic partners.
Brendgen, 2004	Unclear, 37 schools involved but not clear how these were selected; all children from grade 4 to 6 were invited. 61% of the eligible children participated (parental consent).	Yes, exposure and (part of the) outcome measures were self-reported.	No, sex, aggression at T1, depression at T1 were taken into account.	Yes, 20% of the initially included children were excluded because they did not have valid data on their dyadic friendships, and an additional 11% was excluded because of missing data or absence at T2. The remaining participants differed from those who were	Yes, peer-nominations were restricted to the list of children who had received parental consent for participation in the study.

				excluded or lost in that the former were less aggressive, were better accepted by their peers and perceived themselves as more accepted by their peers.	
Buck, 2012	Unclear, study is part of NICHD Study of Early Child Care and Youth Development. No further information on recruitment of participants.	Yes, exposure and outcome measures of interest were self-reported.	No, 11 demographic factors were controlled in all analyses: mothers' marital status, whether fathers lived at home, whether mothers received public assistance, maternal and paternal education level, maternal and paternal employment, income to needs ratio, mothers' depressive symptoms, child ethnicity, and child gender.	No, Full Information Maximum Likelihood (FIML) was used to handle missing data. Of the 1,364 initial participants in the NICHD study, 424 were not available at T2. Analyses were run both using all participants, with FIML to estimate missing data, and with only those tested at Time 2. Results were virtually identical.	N/A
Buist, 2010	No, 23 Dutch municipalities supplied a list of candidate families with at least 2 children between 11 and 15 years old and their 2 parents, all living at the same address. These families received an invitation letter and were contacted by phone.

The response rate was 44/4% | No, outcome (self, sibling, father and mother) and exposure (self and sibling) measures were assessed by multiple informants. | No, multivariate model in which level of sibling relationship quality was related to level and slope of older and younger sibling delinquency was applied. | No, a total of 288 families participated at T1, and 285 at T2 and T3. So, only three families dropped out during the course of the study. | Unclear, the sample consisted of predominantly Dutch siblings from intact middle-class families, showing low levels of delinquent behavior, which may limit the generalizability of the conclusions |
| Cappadocia, 2013 | No, students were selected with a clustered sampling design, using | Yes, outcome and exposure measures were self-reported. | No, binary logistic regression analyses were conducted to investigate | No, reported levels of cyberbullying behaviors at T1 did not differ | |

	class or school as the initial sampling unit, to obtain representative samples of adolescents across Canada. Within each province, samples were also stratified based on school size, location, language, and religion.		whether students with elevated scores on Time 1 covariates were more likely to report increased involvement in cyberbullying/ -victimization over the 1-year period.	significantly between the group of who reported their level of involvement in cyberbullying or cybervictimization at both time points and those who reported level of involvement at T1 only.	
Costello, 2020	Unclear, study is part of The Kids, Lives, Family, and Friends Project. No further information on recruitment of participants.	Yes, outcome measures were self-reported. However, intensity of social interactions were assessed by observers.	No, hierarchical regression models were used to determine how friendship factors operate together in predicting depressive or aggressive symptoms.	No, 94% of the 170 teens who participated at age 16 also participated at age 17. Individuals who did not participate at age 17 did not differ from teens who did on any baseline measures, gender, or familial income.	Yes, the moderating effect in the outcomes tested accounted for a modest amount of variance (β's=0.14 to 0.22). External, unmeasured factors in adolescents' lives could, for example, potentially help further explain changes in symptom levels and close friendship dynamics.
Defoe, 2013	Unclear, the study is part of 'Research on Adolescents Development And Relationships'. No further information on recruitment of participants.	Unclear, exposure and outcomes were self-reported but complimented with multi-informants?	No, cross-lagged paths were used to control for longitudinal links between externalizing problems and negative interactions from/with multi-informants and gender and age.	No, attrition in this study was low. Of the 497 families at T1, 466, 474, and 440 participated at the three follow-up measurements, respectively. Per variable, a maximum of 27% of the cases were missing. Missing data analyses suggested a random pattern of missingness.	Yes, measurement for negative interaction did not make a distinction between constructive and destructive negative interaction.
Elmore, 2010	Yes, participants were recruited from 5 middle schools in 2 districts. The sample was not representative of the US or of	Yes, exposure and outcome measures were self-reported.	No, hierarchical regression models were used. Demographic variables and time 1 levels of school resistance/ aggression	No, at T2 419 of the original 587 students completed the questionnaires. There were no significant differences between those who completed	N/A

86

	local school populations.		behaviour were entered as confounding factors.	the questionnaires at T1 and those at T2.	
Glazebrook, 2015	Unclear, recruitment was through CAMHS staff, it is not known how participants were selected and how many young people were approached who declined to supply contact information. The response rate was 57.1%.	Yes, exposure and outcome measures were self-reported.	No, regression models used and covariates were age, gender, maternal attachment, peer attachment, and levels of previous self-harm at baseline and baseline levels of anxiety and depression.	Unclear, little people were lost during follow-up but consenters were older (median age 15 [IQR = 15–16] vs. 15 [IQR = 14–15]; Z = -2.63, p = .009), although there was no difference in gender between consenters and dissenters.	Yes, males were under represented in the study. Furthermore, as this sample of adolescents contained individuals with elevated levels of depression and anxiety, there is the potential that level of distress impacted attachment classification.
Hazel, 2014	Unclear, letters describing the study were mailed to all parents of children in those grades within participating school districts. It remains unknown how the schools were selected. The response rate was 61%.	Yes, exposure and outcome measures were self-reported.	No, models were used to adjust for multiple covariates such as gender, parental education, grade at baseline, ...	No, the vast majority of youth provided data at each follow up: 91% at 3-months, and then 88% at 6-, 9-, and 12-months and comparisons on available screening variables between those who chose to participate and those who did not, revealed no significant difference on any variable.	N/A
Hedeland, 2016	No, controls were age- and gender-matched	Yes, self-reported, but because the patients were too vulnerable to fill out the questionnaires by themselves, they were filled out by the healthcare staff in cooperation with the patients, based on their mutual conversations about the different questions and the healthcare staffs' background	Yes, controls were matched by age and gender, but no further covariates were taken into account.	No, no dropouts.	Yes, only adolescents who attempted suicide with acetaminophen were included.

		knowledge about the patients.			
Kidd, 2006	No, part of Add Health Study: a clustered sampling design based around 80 high schools and 52 middle schools in the US was selected with unequal probability of selection. Incorporating systematic sampling methods and implicit stratification ensured this sample is representative of U.S. schools with respect to region of country, urbanicity, school type, ethnicity and school size.	Yes, exposure and outcome measures were self-reported	No, controlled for T1 suicide attempts, depressive symptoms, gender and relational domains.	Unclear, no information on loss to follow-up.	N/A
Kingery, 2007	Yes, all schools of a certain region were contacted and self-selected to participate; all children of grade 5 were invited. 68% of the originally contacted children had the consent to participate.	Yes, outcomes were self-reported. Exposure was peer-rated.	No, acceptance, number of friends, friendship quality, loneliness and school involvement were taken into account in the hierarchical regression analysis.	Yes, 13% got lost from T1 to T2, but this was mainly due to moving away. No analyses on the characteristics of drop-outs.	Yes, the fact that the majority of the sample was Caucasian and from low to middle socioeconomic backgrounds may limit the generalizability of these findings to students of diverse racial or ethnic backgrounds. Some children who appeared "friendless" may have had a friend who simply did not have permission to participate in the study. Children were asked to select best friends from a list of their participating

					classmates. As such, a child who had a best friend in another class did not have the option of selecting that individual.
Kingery, 2011	Yes, all schools of a certain region were contacted and self-selected to participate; all children of grade 5 were invited. 62% of the originally contacted children had the consent to participate.	Yes, outcomes were self-reported. Exposure was peer-rated.	No, gender, acceptance, number of friends, friendship quality, loneliness, depression, self-esteem, school involvement, achievement and school avoidance were taken into account in the regression analysis.	Yes, 8% got lost from T1 to T2, but this was mainly due to moving away. No analyses on the characteristics of drop-outs.	See Kingery, 2007
Koch, 2020	Yes, self-selection since recruitment was facilitated through advertisement via emails and canvasing parents at drop-off for summer youth activity programs. The response rate is unknown.	Yes, exposure and outcome measures were self-reported. Single-informant methodologies may be subject to biases that individuals have when reporting about themselves.	No, the path analyses accounted for baseline depressive symptoms in addition to the baseline explanatory variables that have also been linked to depressive symptoms (i.e., pubertal status, rumination, rejection sensitivity, and peer problems). Age was included in all models as a covariate to establish that the effects of pubertal status were independent of the effects of chronological age.	No, there were no significant differences in pubertal development, peer problems, rumination or, anxious or angry rejection sensitivity across participants with complete data at all time points and those missing either T2 role disruption or T3 depressive symptoms. This suggests that the data satisfy the conditions of missing at random, which assumes that the probability of missingness on outcome variables is uncorrelated with the values of the outcome variables themselves.	N/A
Laird, 2013	Yes, no random sampling	Yes, exposure and outcome measures were self-reported	No, multivariable analysis accounted for earlier adjustment, secrecy from	Yes, 18.3% loss of follow-up; girls were more likely to drop out; dropouts had higher quality	N/A

			parents, and the quality of the friendship	friendships at age 12	
Le Grange, 2014	Unclear, authors drew on prospective longitudinal data, which has followed a representative community sample from infancy to adulthood. Further info is lacking.	Yes, exposure and outcome measures were self-reported.	No, multiple known predictor variables were included in the path analysis.	No, although proportionately more families from a lower SES background, or parents who were not born in Australia, have been lost to the study, there were no significant differences on any child characteristics assessed in infancy between the retained cohort at 15–16 years and those no longer participating.	Yes, there was no control for earlier Abnormal Eating Attitudes and Behaviors and hence, although the predictor variables were all antecedent in time to the outcome, it is possible that they could have arisen as consequences of previous Abnormal Eating Attitudes and Behaviors.
McNeil, 2020	No, wide selection across 83 counties using a two-stage combined stratification and cluster design.	Yes, although the use of self-report was necessary to reflect on the child focused approach to maltreatment, it is best to utilize a multimethod approach (e.g., parent and youth) when assessing mental health-related outcomes. The reliance on self-reports in youth may make it difficult to make inferences concerning clinical depression.	No, unconditional parallel process latent growth models were examined for depression, peer relationship problems, and school engagement, before entering them into a single model without predictors. Maltreatment variables and covariates (e.g., gender and race/ethnicity) were then added to the model as predictors of peer relationship problems, school engagement, and depression intercepts (baseline values) and slopes (rates of change).	Unclear, when comparing "completers" and "noncompleters" across all study outcomes. Only T3 depression scores differed across the two groups with noncompleters evidencing higher depression symptoms (p = .020). In other words, those who were present at baseline, missed the T2 follow-up, but then returned for T3 (N = 23; 3.5% of the sample), reported higher depressive symptoms at the last follow-up compared to those who completed all three follow-ups. In order to potentially reduce biases associated with missing data patterns, full-information maximum	Yes, there was an 18-month gap between our follow-up assessments. Using shorter intervals of time (e.g., 3-months) could detect important fluctuations in symptoms that are happening in briefer intervals, especially in early-mid adolescence when greater fluctuations in depressive symptoms may be expected compared to other development epochs.

				likelihood was used to account for missing data.	
Meter, 2015	Yes, participants recruited in only 1 school	Yes, exposure was self-reported. However, the outcome was peer-nominated.	No, stepwise hierarchical regression analysis performed, correcting for T1 aggression, perception of conflict and best friend's aggression at T1.	No, 57/374 did not participate at 2nd timepoint. However, there were no significant differences in age, gender, T1 aggression, or intimate exchange between the individuals who participated at T1 and T2 and those who were not included in the analyses because they did not participate at T2.	Yes, non-participants additionally could not be nominated as best friends in the relationship features peer nomination procedure, which also limited the friends who could be nominated as well as the reciprocal friendships.
Moser, 2002	No, this sample was made possible through access to two recently completed studies of community samples generated by means of random-digit-dailing procedures. The response rate was 51.6%.	Yes, exposure was self-reported. However, the outcome was informed by mother and father separately.	No, regression models to control for outcome at T1, age, gender, parent relationships, sibling relationships and several interaction terms.	No, of the 112 families who participated at T1, 13 families did not participate at T2, due to various reasons such as refusal or an inability to recontact them, resulting in 99 families with complete data. Tests for differences between those families who participated at T2 versus those who did not participate yielded no significant differences with regard to children's and parents' ages, sex of children, and siblings' age differences.	Yes, this study did not consider the effects of peer relationships on the development of deviant adolescent behavior, a particularly important issue given the empirical evidence that peers become increasingly more important to understanding the development of problematic behavior as children enter adolescence.
Munroe, 2020	Yes, participants were recruited via online support networks for caregivers of TGD children and by word of mouth.	No, the parents reported on exposure and outcome of their children.	No, controlled for T1 internalizing and externalizing symptoms	No, no dropouts.	Yes, the racial and financial privilege of the current sample is not reflective of the overall population, and the results of the present study may not be generalizable to the majority of TGD children in the US, who

					may not have the financial ability or support to pursue transition, if it is desired.
Nelemans, 2016	No, participants were recruited from randomly selected schools.\n\nThe baseline response rate was 70%.	Yes, the study exclusively relied on adolescent self-reports.	No, the associations were controlled for longitudinal reverse paths from cannabis frequency to higher peer involvement 1 year later and from peer involvement to lower social anxiety symptoms 1 year later.	Unclear, adolescents participating at all timepoints (85.5%) were slightly younger than those dropping out of the study, but there were no significant differences in gender, or SAD symptoms, peer involvement, and cannabis use at the start of the study.	Yes, cross-lagged panel modeling only allows for inferences about temporal associations.
Nelemans, 2017	Unknown, little info on selection process (recruitment for other project).\nThe baseline response rate was 80%	Yes, the study relied exclusively on youth self-reports of their anxiety symptoms and their social contexts. While youth may be better judges of their own anxiety symptoms than, for example, parents, multi-informant assessments may be particularly important in providing additional information on associations between youth anxiety development and other psychosocial factors.	No, the models were controlled for potential effects of gender, race, and SES on class membership (i.e., gender, race, and SES as predictors of the probability of belonging to the different anxiety trajectory classes) as well as on the initial status and growth factors within the trajectory classes (i.e., gender, race, and SES as predictors of the anxiety level second grade and the subsequent changes in anxiety during elementary and middle school).	Yes, 73.5% of total sample participated at last timepoint. Youth still participating in the study at eighth grade were slightly younger (p=0.03) and more likely to come from low-SES families (p=0.002) than those dropping out. There were no significant differences in gender, race, or anxiety symptoms at the start of the study.	N/A
Nrugham, 2008	Unclear, cluster sampling of schools in 2 counties in Central Norway.	Yes, exposure was self-reported. Outcome was assessed with both	No, outcome was not significant in univariate analysis, and therefore not	Unclear, 2434 of 2464 adolescents participated in both timepoints. No information on differences	N/A

		questionnaire and interview.	included in multivariate analysis.	between those who did and those who did not participate at both times.	
Pace, 2021	No, controls were matched for age and gender.	Yes, outcome and exposure were self-reported.	No, matching for age and gender.	Yes, of the 57 adolescents who were selected because of their BES score, one refused to participate. This was the only boy.	N/A
Patalay, 2018	No, stratified clustered sample design of the original study and to account for sub-group oversampling and attrition over timepoints, all analyses were conducted accounting for the survey design and applying weights. The response rate was 60.9%	Yes, outcome and exposure measures were mostly self-reported. Parents inform peer problems and emotional symptoms.	Yes, regression model which controlled for mental health at age 11 but also e.g. sex, ethnicity, income, …)	Yes, there were substantial levels of attrition in the cohort study and participants were more likely to be missing in the following cases: being male, having a Black ethnicity, lower occupational and educational level of the parents and single parent family. Multiple imputations were carried out to impute values on missing predictors. Overall, missing cells were less than 2% of the total, and 10 imputed datasets were created for analysis.	N/A
Pisarska, 2020	No, random sampling of clusters (class rooms) Response rate was about 83% for T1.	Yes, exposure and outcome measures were self-reported. These data are burdened with recall and social desirability bias.	No, model to control for demographics and T1 gambling involvement	Yes, only 65% response rate from T1 to T2. Youth omitted because of missing data may have on average a higher Gambling Involvement Index score. This selection of students in longitudinal analyses may have biased the sample towards lower risk students in gambling involvement.	N/A
Pittenger, 2018	No, the sample was selected using a two-stage stratified sampling design. In the first stage,	Yes, the use of self-report data may have resulted in underreporting of substance use behaviors.	No, logistic regression models were used to adjust for T1 sociodemograp hic,	No, the sample is weighted to account for sampling strategy and differential selection probabilities	N/A

	states were divided into nine strata and 86 primary sampling units (PSUs), each representing one child welfare service agency (CWS), were randomly selected from these strata. In the second stage, CWS cases were sampled from the PSUs, with oversampling for youth receiving services, those who were in out-of-home placement, and infants. The resulting sample was representative of children and adolescents investigated by and/or receiving services from CWS agencies.		intrapersonal, and interpersonal/ contextual risk and protective factors, and self-reported lifetime use of each respective substance at T1.	resulting from deviation from the sampling strategy during active data collection. These weights also were adjusted for participant non-response.	
Power, 2005	No, randomly selected participants received letters and parental consent forms. However, it is unclear how the schools are selected.	Yes, exposure and outcome measures were self-reported.	No, logistic regression models were used to adjust for other risk and protective factors and control variables were gender, ethnicity and grade in school	Yes, overall, from the first to the fourth time point, approximately 60% of the students remained in the study. T tests on the T1 variables showed that dropouts had higher scores on sensation seeking, mother's attitudes toward adolescent alcohol use, peer involvement in antisocial behavior and hours worked and lower scores on self-esteem, peer security, peer involvement in prosocial activities and school commitment (all p's < .05). Chi-square analyses of T1 drinking patterns showed that dropouts	Yes, the analytic approach had limited power to detect differences because the respondents were divided into a large number of drinking groups, thereby reducing the numbers for statistical comparisons

				were less likely to be abstainers and more likely to be date or heavy, multiple context drinkers (all p's < .05).	
Prinstein, 2005	Yes, little if any info on how selection was performed (school sampling) but probably convenience sampling.

The overall response rate was 74%. | Yes, exposure and outcome measures were self-reported. Of note, friendship quality is also externally informed by the best friend. | No, longitudinal prediction of T3 depressive symptoms by reassurance-seeking, positive friendship quality (self-reported), and gender | Unclear, 13% missing between T1 and T2 and 27% between T1 and T3. Adolescents who participated at both T1 and T2 had higher levels of peer acceptance than adolescents who did not participate at T2. However, attrition analyses revealed no significant differences on any study variables between T2 and T3. | N/A |
| Selfhout, 2009 | No, high schools were randomly selected.

Overall participation rate of the total sample was 42.2% | Yes, exposure and outcome measures were self-reported. The current study assessed the perception of friendship quality by the target adolescent, which may differ from perceptions of the best friends and from objective criteria concerning friendship quality. | No, hierarchical regression modelling in which multiple covariates were added stepwise: sex, outcomes at T1, friendship quality and time spent on internet. | No, the authors compared the 307 participants to the 323 non-participants on depression, social anxiety, and perceived friendship at T1 and T2 and no significant differences (F-values ranging from 0.19 to 0.87, p > 0.10) were found between these groups at either timepoint. | N/A |
| Sund, 2002 | No, schools were drawn with a probability according to size (proportional allocation).

The response rate was 88.3%. | Yes, exposure and outcome measures were self-reported. | No, logistic regression models in which was controlled for depressive level, parent and peer IPPA, stressful events, and demographic variables. | Yes, the attrition rate was only 4.3% but attrition analyses showed that our results could underestimate true relationships. Those who did not participate at T2 were characterized by higher Mood and Feelings Questionnaire (MFQ) total mean scores at T1, lower parent IPPA | Yes, factors not included in the model could have explained the risk associated with low attachment, i.e., hereditary factors, parental psychopathology, school experiences, pubertal maturity, physical |

				scores, and more often having a non-Norwegian background. No peer IPPA, gender, grade, or socioeconomic differences between the two groups were found. Missing data might affect the validity of the results. However, analyses of missing IPPA data did not show any bias.	activity, and somatic heath.
Tu, 2020	Yes, self-selection possible (by parents) since recruitment letters/emails for the study were sent to parents of all 5th grade students at participating elementary schools, as well as distributed in the local community. Parents who responded to the letters/flyers were provided with information about the study.	Unclear, positive peer relationships was evaluated by multi-informants (adolescents, mothers & teachers) but the outcome measure was informed by self-reports. Utilizing various methodologies (e.g., actigraphy, polysomnography, sleep diaries) as well as examining multiple sleep indicators (e.g., bedtime, sleep duration) would provide a stronger assessment of sleep.	No, covariates in the model included adolescent gender, family income, and T1 adolescent reported depressive symptoms.	No, missing data ranged from 3% to 22%, with the latter due to attrition and missing teacher reports at T2. To account for both missingness and non-normality of the data, a robust full wise maximum likelihood estimator (FIML) was used. T tests and chi-square tests across primary study variables revealed no difference between participants who did and did not return at T2.	N/A
Van Zalk, 2015	Unclear, part of a larger study but details on recruitment procedure is lacking.	Yes, the authors have only used adolescents' self-reports, whereas multiple reports would have been preferable. Concerning the measures of friend care, these results may well represent merely the adolescents'	No, models were controlled for the effects of depressive symptoms, self-esteem, gender, age, school, classroom, family structure, several selection effects, such as the network structure and similarity in gender, age,	No, participants who did no fulfill the criteria (and did not have any data on social anxiety or a single friendship nomination at T2) were excluded from the final target sample. They did not, however, differ from the target sample on any of the other study measures at T1, 2, or 3	Yes, adolescents who did not have any friendship nominations were excluded from the sample, yet these may be adolescents with more social anxiety and other related problems.

		own views rather than the actual reality.	and (friends') social anxiety.		
Van Zalk, 2017	Yes, no random sampling	Yes, exposure and outcome measures were self-reported; also at T1-T2 information was collected through a combination of offline and online questionnaires, whereas at T3 all questionnaires were online-only. For online friends all questionnaires were taken online.	No, multivariable analysis accounted for baseline social anxiety, depressive symptoms, friendship quality and friendship stability	Unclear, 72% of the adolescents reported data on all study variables at all three time points, but no attrition analysis	NA
Vaughan, 2010	No, all schools with the targeted grade levels were enrolled and the response rate was 88.4%.	Yes, the exclusive reliance on self-report measures may have inflated observed relationships among study variables due to common method variance and the use of a single informant.	No, all models control for gender, race, parent education, family structure, and cohort.	Yes, comparisons revealed significant differences between study completers and non-completers. Study non-completers endorsed higher levels of depressive symptoms, lower peer support, lower parent education, were more likely to be male and black, and were less likely to be living with both biological parents than study completers.	Yes, authors categorize the factor on friend closeness as peer support but it is more related to quality of relationships
Weerman, 2018	Yes, students from only one school in a rural area included.	Yes, exposure and outcome were self-reported	No, SIENA network analysis performed	Unclear, 155/178 students participated in all 5 timepoints, no attrition analysis reported.	Yes, the current study asks students in retrospect about their time use. This may result in problems with recall, although we believe that this will be relatively minor in because our measurement of unstructured socializing refers to a

					period of only 2 weeks. Measurement of the peer network was limited to 9th grade fellow school students. In reality, adolescents have friends not only from their own school grade, but also from other grades, and more importantly from outside their school.
Wood, 2018	Yes, participants were drawn from 3 public schools from the same city.	Yes, outcome and exposure measures were self-reported.	No, structural equation modelling used.	No, 7.2% missing data at T2, but no significant differences found.	N/A
Zhang, 2018	Yes, participants recruited from only 4 secondary schools in large cities in small municipalities in different areas of the Netherlands. No information if these schools were recruited randomly.	Yes, outcome and exposure measures were self-reported.	No, controlled for gender, age, time-variant components of the relationship predictors at intra- and inter-individual level.	Yes, adolescents who missed participation in one or more timepoints were older, more likely to be boys, and had higher levels of conflicts with parents and peers; but there were no differences in mean levels of satisfaction in the parent-adolescent relationship, satisfaction in the friend-adolescent relationships, or depressive mood.	After completing the survey, adolescents received book certificates for their participation (€5, €7.5, €10, and €12.5 at timepoint 1, 2, 3, and 4, respectively).

Certainty of the body of evidence

5-11 years

Relationships with friends	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Relationships with peers	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data/Large variability of results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Relationships with siblings	Initial grading Low [C]	Downgrading due to
Limitations of study design	0	
Imprecision	-1	Limited sample size/Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

12-18 years

Relationships with friends	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Relationships with peers	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Lack of data/Large variability of results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Relationships with siblings	Initial grading Low [C]	Downgrading due to
Limitations of study design	0	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

5-18 years

Friends/peers/siblings	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data/Large variability of results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	**5-11 years** *Relationships with friends* There is limited evidence <u>with benefit</u> for ***friendship quality***. It was shown that an increase in friendship quality resulted in a statistically significant decrease in loneliness (Kingery 2007) and increase in self-esteem (Kingery 2011). However, a statistically significant change in aggression, depression or school involvement in presence of friendship quality could not be demonstrated (Brendgen 2004, Kingery 2007, Kingery 2011). Of note, there is limited evidence <u>with harm</u> for an overestimation of friendship quality. It was shown that an overestimation of friendship quality resulted in a statistically significant increase in aggression (Brendgen 2004). *Please see the Dutch summary for possible explanations on the latter result.* A statistically significant change in depression when overestimating friendship quality could not be demonstrated. There is limited evidence concerning the risk of psychological outcomes in presence of ***friendship intimacy*** or ***positive relationships with a friend***. A statistically significant change in anxiety or internalizing behavior in presence of friendship intimacy or positive relationships with friends could not be demonstrated (Averdijk 2014, Wood 2019). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data. *Relationships with peers* There is limited evidence <u>with benefit</u> for ***positive peer relationships***. It was shown that positive peer relationships resulted in a statistically significant decrease in sleep/wake problems (Tu 2020). There is limited evidence concerning the risk of psychological outcomes in presence of ***peer problems or conflicts***. A statistically significant increase in internalizing or externalizing symptoms or self-harm when having problems or conflicts with peers could not be demonstrated (Borshman 2020, Munroe 2020). Evidence is of very low certainty and results cannot be considered precise due to limited sample size, lack of data and/or large variability of results. *Relationships with siblings* There is limited evidence <u>with benefit</u> for ***sibling warmth***. It was shown that sibling warmth resulted in a statistically significant increase in emotional and behavioural resilience (Bowes 2010). However, a statistically significant change in internalizing behaviour when having a positive relationship with siblings could not be demonstrated (Averdijk 2014). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data. **12-18 years**

Relationships with friends
There is limited evidence <u>with benefit</u> for ***friendship care, closeness, and satisfaction***.
It was shown that an increase in friendship care resulted in a statistically significant decrease in social anxiety (Van Zalk 2015) and that friendship closeness and satisfaction resulted in a statistically significant decrease in depression-related outcomes (Vaughan 2010, Zhang 2018). However, a statistically significant change in suicidal acts in presence of friend attachment could not be demonstrated (Nrugham 2008).

Depending on the outcome, there is limited evidence <u>with either benefit or harm</u> for ***friendship quality***.
On the one hand, it was shown that an increase in friendship quality resulted in a statistically significant decrease in (social) anxiety (Nelemans 2017, Van Zalk 2017). One other study could not demonstrate a statistically significant change in social anxiety in presence of friendship quality (Selfhout 2009). On the other hand, it was shown that an increase in friendship quality resulted in a statistically significant increase in depressive symptoms and antisocial behaviour (Laird 2013, Van Zalk 2017). *Please see the Dutch summary for possible explanations on this counterintuitive result.*
However, a statistically significant change in depression, cyberbullying and victimization and binge eating in presence of friendship quality could not be demonstrated (Buck 2012, Cappadocia 2013, Laird 2013, Pace 2021, Prinstein 2005, Selfhout 2009).

However, there is also limited evidence <u>with harm</u> for ***friendship intensity***. It was shown that friendship intensity resulted in a statistically significant increase in depressive symptoms (Costello 2020). However, a statistically significant change in depressive symptoms could not be demonstrated in the presence of friendship intensity, when moderated by engagement of close friend (Costello 2020). *Please see the Dutch summary for possible explanations on this counterintuitive result*. It is of note that a statistically significant change in aggression could not be demonstrated in the presence of friendship intensity or with intimate exchange (Costello 2020, Meter 2015).

Moreover, there is limited evidence <u>with harm</u> for ***friend conflict***.
It was shown in a large study (>1000 participants) that conflicts with friend resulted in a statistically significant increase in depressive mood (Zhang 2018). Three other studies could not demonstrate a statistically significant change in depressive symptoms, marijuana and hard drug use, externalizing problems, emotional symptoms or wellbeing when arguing with friends or having negative friend interactions (Branstetter 2011, Defoe 2013, Patalay 2018).

Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data.

Relationships with peers
There is limited evidence <u>with benefit</u> for ***peer involvement and relationships*** for the outcome social anxiety but <u>with harm</u> for the outcomes cannabis use and abnormal eating attitudes. On the one hand, it was shown that peer involvement resulted in a significant decrease in social anxiety (Nelemans 2016) and, on the other hand, peer involvement and strong peer relationships significantly increased cannabis use and abnormal eating attitudes (Nelemans 2016, Le Grange 2014). *Please see the Dutch summary for possible explanations on this counterintuitive result.* Seven other studies could not demonstrate a statistically significant change in depressive symptoms, marijuana or cocaine use, problematic alcohol drinking, gambling, aggression-related symptoms or suicide in the presence of peer attachment, security or when having good peer relationships (Elmore 2010, Kidd 2006, Pisarska 2020, Pittenger 2018, Power 2005, Sund 2002, Weerman 2018).

There is limited evidence <u>with harm</u> for ***having problems with peer relationships or attachment***. It was shown that peer problems and insecure/poor peer attachment resulted in a statistically significant negative change in depressive symptoms, self-harm, emotional symptoms and wellbeing (Patalay 2018, McNeil 2020, Glazebrook 2015).

	Evidence is of very low certainty and results cannot be considered precise due to lack of data and/or large variability of results. *Relationships with siblings* There is limited evidence <u>with harm</u> for **sibling conflict**. It was shown by one (small) study that sibling conflict resulted in a statistically significant increase in internalizing behaviour (Moser 2002). A statistically significant change in internalizing or externalizing problems, delinquent behaviour or binge eating in the presence of different sibling relationships could not be demonstrated in 4 studies (Buist 2010, Defoe 2013, Moser 2002, Pace 2021). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data. **5-18 years** There is limited evidence <u>with benefit</u> for **close relationships with friends or siblings** and <u>with harm</u> for **peer stressors and dissociated relationships with friends and siblings**. It was shown that close relationships with friends or siblings resulted in a statistically significant decrease in suicide attempts in youth (10-17 years) (Hedeland 2016). It was shown that peer stressors led to a statistically significant increase in depressive symptoms in youth aged 7-16 years (Hazel 2014), while another smaller study could not demonstrate a statistically significant decrease in depressive symptoms in youth aged 9-14 years when having problems with peers (Koch 2020). One other study showed that intermediate and dissociated relationships with friends or dissociated relationships with siblings resulted in a statistically significant increase in suicide attempts (Hedeland 2016). A statistically significant change in suicide attempts in the presence of intermediate relationships with siblings could not be demonstrated (Hedeland 2016). Evidence is of very low certainty and results cannot be considered precise due to limited sample size, lack of data and/or large variability of results.
Reference(s)	**Articles** Averdijk M, Eisner MP, Ribeaud D. *Do social relationships protect victimized children against internalizing problems?.* Journal of School Violence 2014, 13(1):80-99. Borschmann R, Mundy LK, Canterford L, Moreno-Betancur M, Moran PA, Allen NB, Viner RM, Degenhardt L, Kosola S, Fedyszyn I, Patton GC. *Self-harm in primary school-aged children: Prospective cohort study.* PLoS One 2020, 15(11):e0242802. Bowes L, Maughan B, Caspi A, Moffitt TE, Arseneault L. *Families promote emotional and behavioural resilience to bullying: evidence of an environmental effect.* J Child Psychol Psychiatry 2010, 51(7):809-817. Branstetter SA, Low S, Furman W. *The influence of parents and friends on adolescent substance use: A multidimensional approach.* J Subst Use 2011, 16(2):150-160. Brendgen M, Vitaro F, Turgeon L, Poulin F, Wanner B. *Is there a dark side of positive illusions? Overestimation of social competence and subsequent adjustment in aggressive and nonaggressive children.* J Abnorm Child Psychol 2004, 32(3):305-320. Buck KA, Dix T. *Can developmental changes in inhibition and peer relationships explain why depressive symptoms increase in early adolescence?* J Youth Adolesc 2012, 41(4):403-413. Buist KL. *Sibling relationship quality and adolescent delinquency: a latent growth curve approach.* J Fam Psychol 2010, 24(4):400-410. Cappadocia MC, Craig WM, Pepler D. Cyberbullying: Prevalence, Stability, and Risk Factors During Adolescence. Canadian Journal of School Psychology 2013, 28(2) 171-192. Costello MA, Narr RA, Tan JS, Allen JP. *The Intensity Effect in Adolescent Close Friendships: Implications for Aggressive and Depressive Symptomatology.* J Res Adolesc 2020, 30(1):158-169. Defoe IN, Keijsers L, Hawk ST, Branje S, Dubas JS, Buist K, Frijns T, van Aken MAG, Koot KM, van Lier PAC, Meeus W. *Siblings versus parents and friends: longitudinal linkages to adolescent externalizing problems.* J Child Psychol Psychiatry 2013, 54(8):881-889. Elmore GM, Huebner ES. *Adolescents' satisfaction with school experiences: Relationships with demographics, attachment relationships, and school engagement behavior.* Psychology in the Schools 2007, 47(6):525-537.

Glazebrook K, Townsend E, Sayal K. *The Role of Attachment Style in Predicting Repetition of Adolescent Self-Harm: A Longitudinal Study.* Suicide Life Threat Behav 2015, 45(6):664-678.

Hazel NA, Oppenheimer CW, Technow JR, Young JF, Hankin BL. *Parent relationship quality buffers against the effect of peer stressors on depressive symptoms from middle childhood to adolescence.* Dev Psychol 2014, 50(8):2115-2123.

Hedeland RL, Teilmann G, Jørgensen MH, Thiesen LR, Andersen J, Study-Associated Pediatric Departments. *Risk factors and characteristics of suicide attempts among 381 suicidal adolescents.* Acta Paediatr 2016, 105(10):1231-1238.

Kidd S, Henrich CC, Brookmeyer KA, Davidson L, King RA, Shahar G. *The social context of adolescent suicide attempts: interactive effects of parent, peer, and school social relations.* Suicide Life Threat Behav 2006, 36(4):386-395.

Kingery JN, Erdley CA. *Peer experiences as predictors of adjustment across the middle school transition.* Education & Treatment of Children 2007, 30(2),73-88.

Kingery JN, Erdley, CA, Marshall KC. *Peer acceptance and friendship as predictors of early adolescents' adjustment across the middle school transition*. Merrill-Palmer Quarterly 2011, 57(3):215–243.

Koch MK, Mendle J, Beam C. *Psychological Distress amid Change: Role Disruption in Girls during the Adolescent Transition.* J Abnorm Child Psychol 2020, 48(9):1211-1222.

Laird RD, Bridges BJ, Marsee MA. *Secrets from friends and parents: longitudinal links with depression and antisocial behavior*. J Adolesc 2013, 36(4):685-693.

Le Grange D, O'Connor M, Hughes EK, Macdonald J, Little K, Olsson CA. *Developmental antecedents of abnormal eating attitudes and behaviors in adolescence.* Int J Eat Disord 2014, 47(7):813-824.

McNeil SL, Andrews AR, Cohen JR. *Emotional Maltreatment and Adolescent Depression: Mediating Mechanisms and Demographic Considerations in a Child Welfare Sample.* Child Dev 2020, 91(5):1681-1697.

Meter DJ, Casper DM, Card NA. *Perceptions of intimacy and friendship reciprocity moderate peer influence on aggression.* Aggress Behav 2015, 41(5):432-442.

Moser RP, Jacob T. *Parental and sibling effects in adolescent outcomes.* Psychol Rep 2002, 91(2):463-479.

Munroe C, Clerkin EM, Kuvalanka KA. *The Impact of Peer and Family Functioning on Transgender and Gender-Diverse Children's Mental Health.* J Child Fam Stud 2020, 29(7):2080-2089.

Nelemans SA, Hale WW, Raaijmakers QAW, Branje SJT, van Lier PAC, Meeus WHJ. *Longitudinal associations between social anxiety symptoms and cannabis use throughout adolescence: the role of peer involvement.* Eur Child Adolesc Psychiatry 2016, 25(5):483-492.

Nelemans SA, Hale WW, Branje SJT, Meeus WHJ, Rudolph KD. *Individual differences in anxiety trajectories from Grades 2 to 8: Impact of the middle school transition.* Dev Psychopathol 2018, 30(4):1487-1501.

Nrugham L, Larsson B, Sund AM. *Predictors of suicidal acts across adolescence: influences of familial, peer and individual factors.* J Affect Disord 2008, 109(1-2):35-45.

Pace CS, Muzi S, Parolin L, Milesi A, Tognasso G, Santona A. *Binge eating attitudes in community adolescent sample and relationships with interview-assessed attachment representations in girls: a multi-center study from North Italy.* Eat Weight Disord 2021, doi: 10.1007/s40519-021-01183-8. Online ahead of print.

Patalay P, Fitzsimons E. *Development and predictors of mental ill-health and wellbeing from childhood to adolescence.* Soc Psychiatry Psychiatr Epidemiol 2018, 53(12):1311-1323.

Pisarska A, Ostaszewski K. *Factors associated with youth gambling: longitudinal study among high school students*. Public Health 2020, 184:33-40.

Pittenger SL, Moore KE, Dworkin ER, Crusto CA, Connell CM. *Risk and protective factors for alcohol, marijuana, and cocaine use among child welfare-involved youth.* Children and Youth Services Review 2018, 95(3):88-94.

Power TG, Stewart CD, Hughes SO, Arbona C. *Predicting patterns of adolescent alcohol use: a longitudinal study.* J Stud Alcohol 2005, 66(1):74-81.

Prinstein MJ, Borelli JL, Cheah CSL, Simon VA, Aikins JW. *Adolescent Girls' Interpersonal Vulnerability to Depressive Symptoms: A Longitudinal Examination of Reassurance-Seeking and Peer Relationships.* Journal of Abnormal Psychology 2005, 114(4), 676-688.

Selfhout MHW, Branje SJT, Delsing M, ter Bogt TFM, Meeus WHJ. *Different types of Internet use, depression, and social anxiety: the role of perceived friendship quality.* J Adolesc 2009, 32(4):819-833.

Sund AM, Wichstrøm L. *Insecure attachment as a risk factor for future depressive symptoms in early adolescence.* J Am Acad Child Adolesc Psychiatry 2002, 41(12):1478-1485.

Tu KM, Cai T. *Reciprocal associations between adolescent peer relationships and sleep.* Sleep Health 2020, 6(6)743-748.

Van Zalk N, Tillfors M. *Co-rumination buffers the link between social anxiety and depressive symptoms in early adolescence.* Child and Adolescent Psychiatry and Mental Health 2017, 11(1).

Van Zalk N, Van Zalk M. *The importance of perceived care and connectedness with friends and parents for adolescent social anxiety.* J Pers 2015, 83(3):346-360.

Vaughan CA, Foshee VA, Ennett ST. *Protective Effects of Maternal and Peer Support on Depressive Symptoms during Adolescence.* J Abnorm Child Psychol 2010, 38(2):261-272.

Weerman F, Wilcox P, Sullivan CJ. *The Short-Term Dynamics of Peers and Delinquent Behavior: An Analysis of Bi-weekly Changes Within a High School Student Network.* Journal of Quantitative Criminology 2018, 34:431-463.

Wood MA, Bukowski WM, Santo JB. *Friendship Security, But Not Friendship Intimacy, Moderates the Stability of Anxiety During Preadolescence.* J Clin Child Adolesc Psychol 2017, 46(6):798-809.

Zhang S, Baams L, van de Bongardt D, Dubas JS. *Intra- and Inter-Individual Differences in Adolescent Depressive Mood: the Role of Relationships with Parents and Friends.* J Abnorm Child Psychol 2018, 46(4):811-824.

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Network of Relationships Inventory (NRI)			
Negative interaction with a friend	Branstetter, 2011	9 items of the Network of Relationships Inventory (NRI)	The Negative Interaction factor was comprised of three, three-item scales: 1) conflict, 2) antagonism, and 3) criticism. Each item was rated using a 5-point Likert scale. See Network of Relationships Inventory.doc
Negative interaction with a sibling or friend	Defoe, 2013	'Negative Interaction' subscale of the Network of Relationships Inventory (NRI)	Negative interactions were assessed with a 5-point Likert scale ranging from 1 (little to none) to 5 (could not be more), and it comprises measures of conflict (three items; e.g., 'How much do you and your sibling disagree and quarrel?) and antagonism (three items; e.g., 'How much do you and your sibling hassle or nag one another?). Thus, higher scores indicate greater quantity (not intensity) of negative interactions. Mean scores across items were used. Reliabilities were acceptable across timepoints. See Network of Relationships Inventory.doc
Friendship intimacy	Wood, 2017	Intimacy subscale of Network Relationship Inventory (NRI)	Intimacy items were "How much do you talk about everything with this person?" "How much do you share your secrets and private feelings with this person?" and "How much do you talk to this person about things that you don't want others to know?" See Network of Relationships Inventory.doc

Friend satisfaction and friend conflict	Zhang, 2018	Satisfaction and conflict subscales of Network Relationship Inventory (NRI)	Adolescents were required to respond based on their relationship with best friends (plural). If the adolescents did not have best friends, then they should base their responses on peers who come closest to that. The scale was comparable to the measure of quality of parent-adolescent relationship, with only the objects of the items replaced by "best friends" (e.g., "How satisfied are you with the relationship with your best friends"). Participants' responses covered the full range of the scale. A higher mean score on the satisfaction subscale indicated higher satisfaction with their friendships and a higher score on the conflict subscale indicated more conflicts with the best friends. See Network of Relationships Inventory.doc
Friendship quality	Prinstein, 2005	7 subscales from the Network of Relationships Inventory (NRI)	Each narrow-band NRI subscale assessed (i.e., Companionship, Criticism, Intimacy, Reliable Alliance, Conflict, Emotional Support, and Dominance) includes three items describing behaviors that occur within the context of the relationship. Adolescents respond to each item using a 5-point Likert scale. Because adolescents in reciprocal best friendships both provided reports of friendship quality on the NRI that pertained to the same relationship, it was possible to use best friends' reports as a measure of friendship quality that relied on an external informant, yielding friend and self-reported measures of friendship quality. See Network of Relationships Inventory.doc
Friendship quality	Selfhout, 2009	The support scale of the Network of Relationship Inventory (NRI)	This scale contained 12 questions, such as: "How often do you turn to your best friend for support with personal problems?". Participants were asked to answer questions about relationship characteristics on a five-point scale (1 = never, 5 = always). The NRI has good predictive, factorial, and construct validity. Internal consistencies of support were 0.89 and 0.91 at T1 and T2, respectively. See Network of Relationships Inventory.doc
Loneliness and Social Dissatisfaction Questionnaire			

Peer relationships	Pittenger, 2018	16-item Loneliness and Social Dissatisfaction Questionnaire (LSDQ)	The LSDQ is designed to measure success in making and keeping friends as well as school adjustment and uses a 5-point Likert scale (1 = never, 2 = hardly ever, 3 = sometimes, 4 = most of the time, 5 = always). Items were recoded so that higher scores indicated greater loneliness and a total score was calculated by summing all 16 items (possible scores range from 16 to 80). This scale demonstrated good to excellent reliability (α = .90). See Cassidy 1992 for more information.
	McNeil, 2020	16-item Loneliness and Social Dissatisfaction Questionnaire for Young Children (LSDQ)	The youth report on the quality of their peer relationships was measured using the Loneliness and Social Dissatisfaction Questionnaire for Young Children at all three time points. This 16-item measure assesses loneliness and social dissatisfaction with peer relationships in young children (e.g., "It's hard to get kids in school to like me"), using a 5-point Likert-type scale (1 = never, 2 = hardly ever, 3 = sometimes, 4 = most of the time, 5 = always). This measure has previously demonstrated strong internal consistency in adolescent populations and exhibited adequate reliability across each timepoint (α = .88, .88, .89).
Inventory of Parent and Peer Attachment (IPPA)			
Peer security	Power, 2005	4 items adapted from IPPA	Questions include e.g., "My friends accept me the way I am". Confirmatory factor analyses yielded excellent fit: χ2 = 2.44, 2 df, p = .29, GFI = 1.00, RMSR = 0.01, standardized factor loadings: 0.68-0.78 (coefficient alpha: 0.81). See Inventory of parent and peer attachment.doc
Peer attachment	Sund, 2002	Revised version of the IPPA	The items are related to three domains of attachment quality: trust items reflect the degree of mutual understanding and respect (e.g., "My mother/father/friends respect my feelings"); communication items assess the extent of spoken communication (e.g., "I tell my mother/father/friends about my problems and troubles"); and alienation items tap feelings of anger and interpersonal isolation (e.g., "My mother/father/friends do not understand what I am going through these days"). The IPPA uses a 5-point Likert-type scale format (1 = "almost never true," 5 = "almost always true"). When the

			total score for each participant was computed, the Alienation subscale items were reverse-scored. High scores reflect greater perceived attachment than do low scores.
Peer attachment	Elmore, 2010 Nrugham, 2008	IPPA	See Inventory of parent and peer attachment.doc
Peer relationships	Le Grange, 2014		
Friendschip quality			
Nelemans, 2017			
Sibling relationship quality	Buist, 2020	The IPPA was designed to measure the quality of communication, trust and (lack of) alienation an individual perceives in a particular relationship. The scale contains 10 items, using a 5-point Likert scale response format (1 = very untrue to 5 = very true). Sample items are: "I easily get upset with my brother/sister" (recoded) and "My brother/sister accepts me as I am." Mean Cronbach's alpha (T1, T2, and T3; older and younger sibling) was .84, range .80 to .88. Older and younger sibling reports were significantly correlated (.41, .44, and .43 for T1, T2, and T3 respectively). For each measurement timepoint, scores of both siblings were combined into one Sibling relationship quality score by computing their mean. Higher scores indicate higher sibling relationship quality.	
Friendship Quality Questionnaire			
Friendship quality	Brendgen, 2004	Short version of the Friendship Quality Questionnaire	After nominating their 5 best friends in school, participants were asked to describe the quality of the friendship with their first nominated (i.e., very best) school friend using a short version (27 items) of the Friendship Quality Questionnaire (FQQ; Parker & Asher, 1993). The items of the FQQ assess six dimensions (Companionship and Recreation, Help and Guidance, Validation and Caring, Intimate Exchange, Conflict Resolution, and Conflict). The

			children were asked to rate how true a specific item description was for their relationship with their best friend, ranging from 0=not at all true to 4=really true. See Parker_1993_Friendship and Friendship quality.pdf
Friendship quality	Buck, 2012	Friendship Quality Questionnaire	This 21-item questionnaire assesses individuals' perceptions of their closest friendship. It uses a 5-point response scale that ranges from 1 ("not at all true") to 5 ("really true") to measure six components of individuals' relationships with their best friend: validation and caring, conflict resolution, conflict and betrayal, help and guidance, companionship and recreation, and intimate exchange. Sample items include, "This friend and I make each other feel important and special," and "This friend and I tell each other private things a lot." A Friendship Quality Total Score was computed by averaging responses across items. Internal consistency for this score was excellent at both Time 1 (α = .93) and Time 2 (α = .92). See Parker_1993_Friendship and Friendship quality.pdf
Friendship quality	Kingery, 2007 Kingery, 2011	Friendship Quality Questionnaire-Revised (FQQ-R)	This questionnaire consists of 40 primary items and one practice item. For each item, children indicated on a 1 (not at all true) to 5 (really true) scale the extent to which a particular quality was characteristic of their relationship with a specific friend (e.g., "_____ makes me feel good about my ideas.", "_____ and I always tell each other about our problems."). Each child completed a FQQ-R questionnaire regarding a particular friend, whose name was inserted into each individual item using word processing software. The FQQ-R is comprised of six subscales (i.e., validation and caring, conflict resolution, conflict and betrayal, help and guidance, companionship and recreation, intimate exchange). An average friendship quality score was used in this study. This measure had high internal consistency both at T1 (α = .95) and at T2 (α = .96), and test-retest reliability of .42 (p < .01). See Parker_1993_Friendship and Friendship quality.pdf
Friendship quality	Van Zalk, 2017	Based on Parker and Asher's scale	Friendship quality with best online friend: Adolescents were asked to think about the very best friend they had nominated (the 1st on their list of nominations). They

			were then asked about the quality of the friendship, as indicated by 6 questions about perceived support and trust based on Parker and Asher's well-used scale. Examples of items were: "My friend supports me when I have an argument with my parents/teachers," "My friend pays attention to my feelings," and "My friend stands by me when others talk about me behind my back." The response items were Not at all true (1), A little true (2), Somewhat true (3), Pretty true (4), and Really true (5)
Friend care	Van Zalk, 2015	Friendship Quality Questionnaire	Participants answered seven questions concerning their perceptions of the quality of the relationship with their first best friend. At the outset of the study, all of the items from the subscale were used, but due to reliability issues in an initial pilot study, two were dropped. Hence, five items were taken from the FQQ, and these were "Sticks up for me if others talk behind my back," "Says 'I'm sorry' if he/she hurts my feelings," "Would like me even if others didn't," "Does not tell others my secrets," and "Cares about my feelings." Two additional items were then added to the final measure, as they were believed to work better in the study setting. The additional items were "Keeps his/her promises" and "Supports me when I have an argument with my parents/teachers." Response items ranged from 1 (Don't agree at all) to 5 (Agree perfectly). The Cronbach's alphas for the current sample were .83 for T1, .85 for T2, and .88 for T3.
Positive peer relationships	Tu, 2020	Compositive measure: - Friendship Quality Questionnaire - Questionnaire on adolescents' friends' prosociality - Checklist of Peer Relations	A composite for positive peer relationships (using standardized scores) included adolescent-reported friendship quality; adolescent-, mother-, and teacher-reported friends' prosociality; and mother- and teacher-reported peer acceptance. - Adolescents completed the Friendship Quality Questionnaire (25 items). Items (e.g., My friends... "care about my feelings," "make me feel good about my ideas") were rated on a 5-point scale (0 = not at all true to 4 = really true). See Parker_1993_Friendship and Friendship quality.pdf - Adolescents, mothers, and teachers reported on adolescents'

			friends' prosociality (e.g., "make good grades," "have good ideas about fun things to do"). Items were rated on a 5-point scale (1 = never do this to 5 = very often; 7-9 items). - Mothers and teachers completed the 6-item Checklist of Peer Relations to report on adolescents' peer acceptance (e.g., "This child is accepted by peer group," "Other children like this child and seek him or her out."). Items were rated on a 5-point scale (1 = never true to 5 = almost always true). Across both timepoints and across all informants, significant correlations ranged from 0.23 to 0.65 (Ps < 0.05). At T2, adolescent-reported friends' prosociality was not correlated with mother-reported peer acceptance (r = 0.19, p = .08). All measures demonstrated adequate to high reliability across T1 and T2 (as range from 0.67 to 0.95).
Other scales			
Friendship quality	Laird, 2013	Shortened Friendship Quality Scale (FQS)	The quality of the participants' current best friendship at ages 12 and 13 was reported using selected items from the Friendship Quality Scale (FQS). To minimize the length of the interview, a shortened version of the FQS (i.e., three highest loading items from each of the Help, Security, and Closeness subscales was used to assess the quality of the friendship e.g., "If other kids were bothering me, my friend would help me"). Adolescents responded to the questions about their current best friend on a five-point scale (1 = never to 5 = always). An index of friendship quality was computed as the mean of the nine items at ages 12 and 13.
Friendship intensity	Costello, 2020	Intensity of observed social interactions with peers, assessed with: - Autonomous-Relatedness Coding System for Adolescent Peer Dyads - Supportive Behavior Task Coding System for Adolescent Peer Dyads	- Target teens and their close friends participated in two observed social interactions in private offices within a university building. The first observational, "revealed differences" task required teens to discuss which individuals from a list of fictional characters they should vote off of an island in a hypothetical reality TV show, based on the descriptions provided. They were filmed for 8 minutes after being instructed to come to a consensus about which people should be voted off the island. These interactions were coded using the Autonomy-Relatedness Coding System for Adolescent Peer

			Dyads. Autonomous-Relatedness captures the degree to which the target teen states reasons in supporting their position while still maintaining positive connection in the relationship. Indicators of connection include the target teen's physical and verbal demonstrations of warmth and an absence of undermining statements as the teen makes his or her case. Autonomous-Relatedness is coded on a scale from 0–4. Scores of 0 indicates that the participant was unable to state reasons for disagreeing with their friend and did not appear confident, while also lacking in behaviors that promote connectedness between the dyad, such as employing information-seeking queries, offering validation or agreement, and appearing engaged in the task. Coders assign scores of 4 when a participant is able to employ behaviors that both offer support for their position in the disagreement while also maintaining a warm, connected relationship throughout. - The target teen asking their close peer for advice on a self-selected topic. This interaction was then coded for the close peer's engagement using the Supportive Behavior Task Coding System for Adolescent Peer Dyads. Engagement codes range from 0, in which a supporter provides little or no physical or verbal indication of listening attentively or with interest (e.g., closed body posture, restricted eye contact, ignoring the support-seeker), to 4, which indicates that the supporter displays high investment and responsiveness to the support-seeker, both verbally and non-verbally (e.g., consistent eye contact, open posture, verbal follow-ups on the support-seeking statements).
Peer stressors	Hazel, 2014	8 items of the Adolescent Life Events Questionnaire, ALEQ	Items identified a priori as having to do with peer relationships (e.g., "Feeling pressure by friends", "Fighting with or problems with a friend", "Friend is criticizing you behind your back") were summed to form a scale of peer stressor exposure. Each item was rated on a scale from 0 ("never") to 4 ("always") reflecting how often that experience had happened to the participant in the last 3 months. Scores hence ranged from 0 to 32.

			See Hankin 2002 for more info (limited info on questionnaire available).
Peer problems	Koch, 2020	25-item of The Index of Peer Relations, IPR	The Index of Peer Relations was used to assess peer problems at T1 (IPR). The IPR is a 25-item measure designed to assess the severity of problems in peer relationships and frequency of peer conflict. Each item is scored on a 7-point scale where 1 = none of the time and 7 = all of the time. Items were modified to ask about "kids my age" rather than "my peers." For instance, the item "I get along very well with my peers" was modified to "I get along very well with kids my age." Total scores are calculated by taking the sum score of all items and subtracting from this value the number of total items answered. This value is then multiplied by 100 and divided by the product of total items answered multiplied by six. Total scores range from 0 to 100 where higher scores indicate greater problems with peers. A score of 30 or greater indicates a clinically relevant threshold of peer problems. Scores in this sample at T1 ranged from 0 to 84.67 (M= 27.62, SD =16.90). Internal consistency in this sample was α = 0.96 at T1. See IPR.pdf for more info.
Peer attachment	Glazebrook, 2015	The Attachment Questionnaire for Children (AQC)	This questionnaire consists of three descriptions relating to relationships with close friends. Respondents endorse the description that matches their peer relationships most closely giving classifications of secure, insecure-avoidant, or insecure- ambivalent peer attachment. This brief measure has demonstrated good concurrent validity with the Inventory of Parent and Peer Attachment (see above). See The Attachment Questionnaire for Children.docx for more information. See Inventory of parent and peer attachment.doc
Peer problems	Patalay, 2018	Peer problems subscale of the Strengths and Difficulties Questionnaire	See Goodman_1997 for more information.
Peer relations, sibling relations	Pace, 2021	Friends and family interview (FFI)	The FFI is a semi-structured interview asking adolescents (aged 11–17) a set of 27 questions about themselves and their relationships with the most significant individuals in their lives, including parents, best friends, siblings and favorite teacher. FFIs were

			videotaped and transcribed verbatim and transcripts were rated by expert and reliable coders (first and second authors) according to the FFI coding system. Based on the highest score, it is possible to highlight one of the four attachment classifications: secure, insecure-dismissing, insecure-preoccupied, insecure-disorganized. The FFI coding system also includes the following dimensional scales: (1) coherence, based on Grice's maxims of good conversation, such as truth, economy, relation, and manner, plus overall coherence; (2) reflective functioning, in terms of developmental perspective, theory of mind, and diversity of feelings; (3) evidence of secure base (father, mother, and other significant figure); (4) evidence of self–esteem, in terms of social competence, school competence, and selfregard; (5) peer relations, intended as frequency of contact and quality of best friendship; (6) sibling relations, in terms of warmth, hostility, and rivalry; (7) affective regulation in attachment relationships, intended as idealization, role reversal, anger, derogation, and adaptive response; and (8) differentiation of parental representations.
Sibling warmth and control	Moser, 2011	Sibling Relationship Questionnaire	This questionnaire is designed to assess qualitative aspects of the sibling relationship and has 16 scales. These scales assess domains such as Maternal Partiality (e.g., "Why usually gets treated better by your mother, you or this sibling?"), Affection ("How much do you and this sibling love each other?"), Competition ("Some siblings try to out-do or beat each other at things a lot, while other siblings try to out-do each other a little. How much do you and this sibling try to out-do each other at things?"), and Dominance of Sibling ("How much do you tell this sibling what to do?"). A previous principal components analysis by the authors of the questionnaire identified four factors of Warmth, Rivalry, Status/Power, and Conflict. The Warmth factor consists of scale scores for intimacy, admiration by sibling, and affection (15 items); Rivalry consists of maternal and paternal partiality (6 items); Status/Power consists of nurturance of sibling and dominance over sibling, minus the

			scale scores of nurturance by sibling and dominance by sibling (12 items total); and Conflict scores consist of quarreling, antagonism, and competition (6 items). Subjects answered questions regarding these domains using 5-point rating scales with anchors of 1: Hardly at all and 5: Extremely Much. Previous test-retest scale reliabilities performed by the authors yielded values ranging from .58 to .86 (M = .71). See Furman_1985 and Dalton_2016 (cfr. Appendix 3 for full questionnaire) for more information.
Friend closeness	Vaughan, 2010	Question on how close they felt to each friend	Participants provided the names of as many as five of their closest friends at every timepoint and indicated how close they felt to each friend on a scale of 0 (Not close at all) to 3 (Very close). At every timepoint, youths' ratings of closeness to each friend were averaged to derive a composite scale score. Friendship closeness appears to correspond closely to the "intimacy" aspect of friendships captured in other widely used measures of peer relations such as the Network of Relationships Inventory (see above). Although participants could list fewer than five friends, the great majority, i.e., between 83 and 89% for T1–5, reported their level of closeness to five friends. Good coherence of individuals' ratings of closeness to each friend at every timepoint was suggested collectively by estimates of Cronbach's alpha, which ranged from 0.72 to 0.80 across the five timepoints.
Positive relationship with friends	Averdijk 2014	Social Behaviour Questionnaire (SBQ)	See Social Behavior Questionnaire.pdf
Peer involvement	Nelemans, 2016	5-item "intensity of contact with friends" subscale of the Questionnaire on Peer Relationships	Adolescents rated the items on a 3-point scale, ranging from 1 (indicating low peer involvement) to 3 (indicating high peer involvement). Sample items included "How much time do you spend with your peers on weekdays after school time?" and "How often do you meet with your peers in the weekend (Saturday and Sunday)?". In this study, reliability of the peer involvement subscale was acceptable over all timepoints ($α = 0.66–0.72$).

Thematic category: Being loved and being part of the group

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Barzeva, 2020, The Netherlands	Observational: prospective cohort study (3 timepoints in a a 5-year period)	2772 adolescents, aged 11-16 years (mean age at T1 11.11±0.55 years; at T3 16.21±0.72 years); the cohort was a combination of children from 5 municipalities and all schools from the North of the Netherlands, and a clinically referred cohort (2 child psychiatric outpatient clinics).	Peer acceptance (four items of the Classmate affection subscale and four items of the Classmate behavioral affirmation subscale of the self-reported Social Production Function) (see below for more details)	Outcomes measured: - Social anxiety (Social phobia subscale of the Revised Children's Anxiety and Depression Scale) - Social withdrawal (Youth Self-reported social withdrawal subscale, parent-reported Child Behavioral Checklist depressive/withdrawn scale) Statistics: random intercept cross-lagged panel model
Brendgen, 2004, Canada	Observational: prospective cohort study (2 timepoints in a 6-month period)	819 children, aged 8 to 13 years (mean age at T1 10.3 years, SD not available); all children were from grade 4-6 from 37 schools in low to average SES areas in Montreal.	Social acceptance (social acceptance subscale of the Self-Perception Profile for Children, self-perceived) (see below for more details)	Outcomes measured: - Aggression (five peer-rated items taken from the Pupil Evaluation Inventory) - Depression (Children's Depression Inventory, self-reported) Statistics: Hierarchical Multiple Linear Regression Analysis
Grills-Taquechel, 2010, USA	Observational: prospective cohort study (2 timepoints in a 2-year period)	77 children, aged 11 to 14 years (mean age at T1 11.69±0.52 years, at T2 13.64±0.54 years). Participants were students enrolled in a southwestern Virginia middle school in Grade 6.	Social acceptance (social acceptance subscale of Self-perception Profile for Children, self-report) (see below for more details)	Outcomes measured: Anxiety (Social anxiety subscale of Multidimensional Anxiety Scale for Children, self-report) Statistics: Multiple Regression Analyses
Hughes, 2001, USA	Observational: prospective cohort study (2 timepoints in a an 18-month period)	49 children, aged 7-8 years (grade 2 and 3) (mean age at T1 7.63±0.78 years). Children were nominated by 44 2nd and 3rd grade teachers	Peer acceptance (subscale of the Pictorial Scale of Perceived Competence and Social Acceptance for Young Children) (see below for more details)	Outcomes measured: Aggression (Peer nominations of relational and overt aggression were obtained using procedures described by Masten, Morrison,

		from 5 local elementary schools.		and Pelligrini (1985); peer-rated)
				This study was part of an intervention study but treatment (2 mentoring interventions) was not linked to the risk factor and outcome, which makes it unlikely that this affects the pattern of relations in the study. Statistics: hierarchical regression analyses
Kingery, 2007, USA	Observational: prospective cohort study (2 timepoints in a 6-month period)	146 children, aged 10-11 (grade 5) (mean age at T1 11 years and 1 month). Elementary and middle schools from two public school districts located in low- to middle-income communities in northern New England were recruited.	Peer acceptance (peer-rated score based on the question "how much do you like to play with this person at school?")	Outcomes measured: - Loneliness (Loneliness and Social Dissatisfaction Questionnaire, self-reported) - School involvement (Attitudes Toward School self-reported questionnaire) Statistics: hierarchical regression analyses
Kingery, 2011, USA	Observational: prospective cohort study (2 timepoints in a 6-month period)	365 adolescents, aged 10-11 (grade 5) (mean age at T1 11 years and 2 months). Elementary and middle schools from six public school districts located in lower- to middle-income rural and suburban communities in northern New England were recruited.	Peer acceptance (peer-rated score based on the question "how much do you like to play with this person at school?")	Outcomes measured: - Loneliness (Loneliness and Social Dissatisfaction Questionnaire, self-reported) - Depression (the Children's Depression Inventory, self-reported) - School involvement (Attitudes Toward School self-reported questionnaire) - Self-concept (Self-Perception Profile for Children) Statistics: regression analyses
Klima, 2008, USA	Observational: prospective cohort study (2 time-points in a 2-year period)	156 children, aged 9-12 years (grade 4-6) (mean age at T1 9.5 years). The children were 4th graders at one parochial school and two public schools in a large metropolitan area.	Peer acceptance (8-item measure, developed by the authors, teacher-rated, based on how well accepted a child is by classmates by questioning if a child is (dis)liked and the extent to which it is excluded)	Outcomes measured: - Self-worth (Global Self-Worth subscale of the Self-Perception Profile for Children, self-reported) - Depression (Children's Depression Inventory, self-reported)

				- Internalizing symptoms (teacher report form) - Externalizing symptoms (teacher report form) Statistics: Hierarchical Multiple Regression Analyses
McDonough, 2016, New Zealand	Observational: prospective cohort study (3 time-points in a 3 year period)	1940 adolescents, aged 10-15 years (mean age at T1 12.20±1.75 years). Participants were recruited from 78 schools on the North Island of New Zealand.	Peer connectedness (seven items were generated for the present study to assess relationships with peers at school, happiness with number of close friends, and support from friends, self-reported)	Outcomes measured: - Marijuana use (based on the Washington Healthy Youth Survey, self-reported) - Other illegal drug use (based on the Washington Healthy Youth Survey, self-reported) Statistics: cross-lagged panel models
Selby, 2013, USA	Observational: prospective cohort study (weekly ratings in a 6-month period)	119 adolescents after a psychiatric hospitalization, aged 12 to 18 years (mean age 15.3±1.4 years). Participants were recruited from the adolescent inpatient unit of a psychiatric hospital in the Northeast.	Peer invalidation (based on the following questions: Were there times when you did not feel accepted by your classmates? Or that you were being left out? Or that you could not express your true thoughts or feelings? Or that if you did express your thoughts and feelings, that you would be dismissed, punished, ignored, or made fun of? How many friends do you have that can confide in? Self-reported via interview) Weekly ratings were mapped via recall and with the help of time anchors during the 6-month follow-up interview.	Outcomes measured: Suicidal ideation (based on the following questions: Sometimes when people are upset or feel bad they think about dying or even killing themselves. Do you have these thoughts during these past six months? How often did you have these thoughts? When you had them, how long did they last – a few minutes, an hour or more, almost always? Did you have these thoughts throughout the past month or where there times when you did not have these thoughts? Do you have those thoughts now? Do you have a plan? What is it? Have you told anyone about these thoughts or plans? Self-reported via interview) Statistics: hierarchical model with lag variables
Teachman, 2007, USA	Observational: prospective cohort study (3 time-points in a 3-year period)	185 adolescents, aged 13 to 18 (mean age at T1 13.35±0.64 years). Adolescents were	Perceived social acceptance (Social acceptance subscale of the Self-Perception Profile for Adolescents)	Outcomes measured: social anxiety (Social Anxiety Scale for Adolescents, self-reported)

		recruited from the 7th and 8th grades at a single public middle school drawing from suburban and urban populations in the South-eastern United States.	(see below for more details)	Statistics: structural equation modeling
Tetzner, 2017, Germany	Observational: prospective cohort study (3 time-points in a 4-year period)	7977 adolescents (age range not available; mean age at T1 13.5±0.68 years). Students were chosen after randomly selecting schools and then taking two full classes (7th grade) per school.	Peer acceptance (based on 3 items: "When the others do something together at recess, they often don't include me"; "No matter what I do, my classmates don't like me"; "Sometimes I feel like an outsider in my class", self-reported)	Outcomes measured: Self-esteem (4-item German version of the Rosenberg Self-esteem Scale, self-reported) Statistics: Cross-lagged panel and multi-group models
Van Voorhees, 2008, USA	Observational: prospective cohort study (2 time-points in a 1-year period)	4791 adolescents, aged 12-18 (grade 7-12) (mean age not available), drawn from the AddHealth study which included 80 high schools	Peer acceptance (participants were asked if they felt socially accepted, self-reported)	Outcomes measured: Depression (depression-specific items from the Center for Epidemiologic Studies Depression scale, self-reported) Statistics: logistic regression analyses
Vanhalst, 2013, Belgium	Observational: prospective cohort study (3 time-points in 2 year period)	882 Belgian adolescents (age range not available; mean age at T1 14.95±0.94 years); all 9th and 10th grade students from three schools in Belgium were invited.	Perceived social acceptance (5-item Perceived Social Acceptance scale of the Adolescent Self-Perception Profile, self-reported) (see below for more details) ["Actual social acceptance was not extracted because it was based on peer nominations]	Outcomes measured: Loneliness (of the Loneliness and Aloneness Scale for Children and Adolescents, self-reported) Statistics: cross-lagged path analysis
Wagner, 2018, Germany	Observational: prospective cohort study (4 time-points (younger cohort) or 2 timepoints (older cohort) with a 1-year interval)	1535 adolescents: 846 grade 5 students (subset of a sample of 2281 students, mean age at T1 10.2±0.65 years) and 689 grade 8 students (subset of a sample of 1766 students, mean age at T1 13.3±0.9 years) (age range not available)	Social inclusion (social self-concept questionnaire, based on 4 items: ("When my classmates do something together during recess, they often ignore me"; "My classmates quite like me"; "No matter what I do, my classmates don't like me"; "Sometimes I feel a little like an outsider in my class", self-reported))	Outcomes measured: Self-esteem (KINDL-R Self-esteem subscale, self-reported) Statistics: fitted two-level multilevel models
Webb, 2016, Australia	Observational: prospective cohort study (2 time-points in a	367 adolescents, aged 9-14 years (mean age	Peer acceptance (based on 1 question: "How much do you feel that	Outcomes measured: Body dismorphic disorder symptoms (Appearance Anxiety

	a 1-year period)	12.1±0.91 years) from grades 5 to 7.	other kids in your school like you?", self-reported)	Inventory, self-reported) Statistics: cross-lagged path model
Yen, 2015, USA	Observational: prospective cohort study (2 timepoints in a a 6-month period)	119 adolescents, from an adolescent inpatient psychiatric unit on the basis of having been recently admitted to the unit for elevated suicide risk, age unknown	Peer invalidation (based on these questions: "Were there times when you did not feel accepted by your classmates? Or that you were being left out? Or that you could not express your true thoughts or feelings? Or that if you did express your thoughts and feelings, that you would be dismissed, punished, ignored, or made fun of? How many friends do you have that can confide in?", self-reported) (weekly ratings)	Outcomes measured: -Suicide events (a suicide attempt or emergency intervention to intercede an attempt, self-reported), -Self-harm (Functional Assessment of Self-Mutilation, self-reported) Statistics: multivariate logistic regression

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
5-11 years				
Peer acceptance				
Self-esteem	Social inclusion (within-person) (10 years old)	Statistically significant: B: 0.27±0.03 £ (p<0.05) *With benefit for social inclusion*	1, 846	Wagner, 2018
	Peer acceptance	Not statistically significant: B: 0.08±0.05 £† (p>0.05)	1, 365 §	Kingery, 2011
Aggression (teacher-rated)		Not statistically significant: β:0.06 £† (p>0.05)	1, 49 §	Hughes, 2001
Aggression (peer-rated)		Not statistically significant: β:0.08 £† (p>0.05)		
Loneliness		Not statistically significant: B: -0.33±0.10 £† (p>0.05) (remark: β is statistically significant, and study authors base their conclusions on the β values)	1, 146 §	Kingery, 2007
		Not statistically significant: B: -0.12±0.05 £† (p>0.05) (remark: β is statistically significant, and study authors base their conclusions on the β values)	1, 365 §	Kingery, 2011
School involvement		Not statistically significant: B: 0.22±0.09 £† (p>0.05) (remark: β is statistically significant, and study authors	1, 146 §	Kingery, 2007

		base their conclusions on the β values)		
		Not statistically significant: B: 0.14±0.06 £† (p>0.05) (remark: β is statistically significant, and study authors base their conclusions on the β values)	1, 365 §	Kingery, 2011
Self-worth		Not statistically significant: B: -0.01±0.01 £† (p>0.05)	1, 156 §	Klima, 2008
Depression		Statistically significant: B: -0.09±0.03 £ (p<0.05) *With benefit for peer acceptance*		
		Not statistically significant: B: -0.90±0.59 £† (p>0.05)	1, 365 §	Kingery, 2011
Internalizing symptoms		Statistically significant: B: -0.04±0.01 £ (p<0.05) *With benefit for peer acceptance*	1, 156 §	Klima, 2008
Externalizing symptoms		Statistically significant: B: -0.04±0.01 £ (p<0.05) *With benefit for peer acceptance*		

12-18 years

Peer acceptance

Self-esteem	Social inclusion (within-person) (13 years old)	Statistically significant: B: 0.13±0.06 £ (p<0.05) *With benefit for social inclusion*	1, 689	Wagner, 2018
Self-esteem (T1 --> T2)	Peer acceptance	Not statistically significant: B: 0.04, 95%CI [-0.06;0.15] £† (p=0.392)	1, 7977	Tetzner, 2017
Self-esteem (T2 --> T3)		Statistically significant: B: 0.08, 95%CI [0.04;0.13] (p<0.001) *With benefit for peer acceptance*		
Depression		Statistically significant: Corrected for demographic variables: OR: 0.04, 95%CI [0.01;0.18] Corrected for baseline depression: OR: 0.2, 95%CI [0.05;0.81] (p<0.05) *With benefit for peer acceptance*	1, 4791	Van Voorhees, 2008
Social anxiety (T1 --> T2)		Not statistically significant: ££† (p>0.05)	1, 2772	Barzeva, 2020
Social anxiety (T2 --> T3)		Not statistically significant: ££† (p>0.05)		
Social withdrawal (T1 --> T2)		*Self-report:* Statistically significant: β: -0.09±0.03 £ (p<0.05) *With benefit for peer acceptance* *Parent-report:* Not statistically significant: £† (p>0.05)		

Social withdrawal (T2 --> T3)		*Self-report:* Not statistically significant: £† (p>0.05) *Parent report:* Statistically significant: β: -0.10±0.04 £ (p<0.05) *With benefit for peer acceptance*		
Social acceptance/social competence				
Social anxiety	Perceived social acceptance	Statistically significant: β: -0.60 £ (p<0.05) *With benefit for perceived social acceptance*	1, 185 §	Teachman, 2007
Anxiety		Statistically significant: β: -0.30 £ (p=0.02) *With benefit for perceived social acceptance*	1, 77 §	Grills-Taquechel, 2010
Loneliness (T1 --> T2)		Statistically significant: β: -0.12 £ (p<0.01) *With benefit for perceived social acceptance*	1, 882	Vanhalst, 2013
Loneliness (T2 --> T3)		Statistically significant: β: -0.12 £ (p<0.01) *With benefit for perceived social acceptance*		
Peer connectedness				
Marijuana use	Peer connectedness	Not statistically significant: ££† (p>0.05)	1, 1940	McDonough, 2016
Other illegal drug use		Not statistically significant: ££† (p>0.05)		
Peer invalidation				
Suicidal ideation (future weekly level)	Peer invalidation	Statistically significant: β: 0.11±0.03 £ (p<0.001) *With harm for peer invalidation*	1, 119 §	Selby, 2013
Suicide events (boys)		Not statistically significant: ££ OR: 1.54, 95%CI [0.57;4.14] ¥ (p>0.05)	1, 119 §	Yen, 2015
Self-harm (boys)		Not statistically significant: ££ OR: 2.45, 95%CI [0.99;6.10] ¥ (p>0.05)		
Suicide events (girls)		Not statistically significant: ££ OR: 1.04, 95%CI [0.58;1.86] ¥ (p>0.05)		
Self-harm (girls)		Statistically significant: ££ OR: 1.86, 95%CI [1.07;3.24] (p<0.05) *With harm for peer invalidation*		
5-18 years				
Peer acceptance				
Body dysmorphic disorder symptoms	Peer acceptance	Not statistically significant: ££†	1, 1367	Webb, 2016

		(p>0.05)		
Social acceptance/social competence				
Aggression	Perceived social acceptance	Statistically significant: B: 0.07 £ (p<0.05) *With benefit for social acceptance*	1, 819	Brendgen, 2004
Depression		Statistically significant: B: -0.11 £ (p<0.05) *With benefit for social acceptance*		

B/β ± SE (unless otherwise indicated), OR: odds ratio, SE: standard error, B: unstandardized coefficient, β: standardized coefficient

£ No SE and/or CI available; or no information on magnitude of effect available to assess variability of results.

££ No raw data or effect size available

¥ Imprecision (large variability of results)

† Imprecision (lack of data)

§ Imprecision (limited sample size or low number of events)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Barzeva, 2020	Unclear, combination of population-based survey (5 municipalities and all schools in a certain region were contacted) and a clinically referred cohort (2 child psychiatric outpatient clinics involved). Response rate at T1 was 76% for the population-based survey; unclear participation rate for the second cohort.	Yes, exposure and (part of the) outcome measures were self-reported.	No, both outcomes (social anxiety, social withdrawal) are taken into account in the path analysis (but not controlled for baseline).	Unclear, T2 and T3 included 81%–96% of the T1 participants, but not clear if there was disproportional loss to follow-up.	N/A
Brendgen, 2004	Unclear, 37 schools involved but not clear how these were selected; all children from grade 4 to 6 were invited. 61% of the eligible children participated	Yes, exposure and (part of the) outcome measures were self-reported.	No, both outcomes, social preference and social acceptance were taken into account in the multiple linear regression analysis, and outcomes were	Yes, 20% of the initially included children were excluded because they did not have valid data on their dyadic friendships, and an additional 11% was excluded	Peer-nominations were restricted to the list of children who had received parental consent for participation in the study.

	(parental consent).		controlled for baseline values.	because of missing data or absence at T2. The remaining participants differed from those who were excluded or lost in that the former were less aggressive, were better accepted by their peers and perceived themselves as more accepted by their peers.	
Grills-Taquechel, 2010	Yes, one specific school was involved; all children of grade 6 were invited and included (parental consent).	Yes, exposure and outcome measures were self-reported.	No, self-worth, social acceptance, social support and gender were taken into account in the multiple regression analysis.	No, only 33% of the initial participants were retained at T2 because of parental consent at T2, but no significant differences were found on any of the demographic or primary variables.	N/A
Hughes, 2001	Yes, 5 specific schools were involved, and teachers nominated children to participate. 75% of the selected children participated in the study (parental consent).	Yes, exposure measure was self-reported.	No, baseline measures, ethnicity and gender taken into account in the multiple regression analysis.	Unclear, T2 included 79% of the T1 participants (62 children). Not clear if drop-out happened disproportionally.	Because subjects were involved in an intervention study, it is possible that treatment may have altered the relations between perceived peer acceptance and subsequent social preference. However, treatment condition did not differentially relate to social preference or aggression, nor did it interact with perceived peer acceptance in predicting these variables. Therefore, it is unlikely treatment affected the pattern of relations in this study, according to the study authors.

Kingery, 2007	Yes, all schools of a certain region were contacted and self-selected to participate; all children of grade 5 were invited. 68% of the originally contacted children had the consent to participate.	Yes, outcomes were self-reported. Exposure was peer-rated.	No, acceptance, number of friends, friendship quality, loneliness and school involvement were taken into account in the hierarchical regression analysis.	Yes, 13% got lost from T1 to T2, but this was mainly due to moving away. No analyses on the characteristics of drop-outs.	- The fact that the majority of the sample was Caucasian and from low to middle socioeconomic backgrounds may limit the generalizability of these findings to students of diverse racial or ethnic backgrounds. - Some children who appeared "friendless" may have had a friend who simply did not have permission to participate in the study. - Children were asked to select best friends from a list of their participating classmates. As such, a child who had a best friend in another class did not have the option of selecting that individual.
Kingery, 2011	Yes, all schools of a certain region were contacted and self-selected to participate; all children of grade 5 were invited. 62% of the originally contacted children had the consent to participate.	Yes, outcomes were self-reported. Exposure was peer-rated.	No, gender, acceptance, number of friends, friendship quality, loneliness, depression, self-esteem, school involvement, achievement and school avoidance were taken into account in the regression analysis.	Yes, 8% got lost from T1 to T2, but this was mainly due to moving away. No analyses on the characteristics of drop-outs.	See Kingery, 2007
Klima, 2008	Yes, convenience sample of 3 schools; all children of grade 4 were invited. 37% of the eligible children participated	Yes, two of the outcomes were self-reported. Exposure was teacher reported.	No, baseline outcomes measures were taken into account in the hierarchical multiple regression analysis.	No, at T2 92% of the T1 children were retained, but only 69% of the children had teacher-data. In comparing children who did and did not have teacher data at T2, no	In exchange for participation, the children received a $5—$10 honorarium at each time point. Teachers completed ratings for participating children in their

				differences were found on any of the T1 variables.	classroom and received $5 for each completed child questionnaire.
McDonough, 2016	Unclear, 78 schools involved but not clear how these were selected; unclear how participants from the schools were recruited.	Yes, exposure and outcomes were self-reported.	No, baseline outcomes and peer connectedness and peer influence were taken into account in the cross-lagged panel model.	No, data from youth who participated in at least two time points were used in the current study (10.8 % attrition from T1). The final sample was nationally representative for gender and SES.	According to the authors, other potential moderators should be explored. The statistical model contains some unexplained variance.
Selby, 2013	No, a specific psychiatric hospital in the Northeast was included; all participants fulfilling the selection criteria and who had consent from their parents were included.	Yes, exposure and outcome measures were self-reported.	No, baseline outcomes and a range of other factors were included in the model.	No, no loss to follow-up	N/A
Teachman, 2007	Yes, adolescents were recruited from the 7th and 8th grades at a single public middle school. Not clear how participants were selected: one cohort of 8th graders was included, and 2 different cohorts of 7th graders were included. Of all students eligible for participation 63% agreed to participate.	Yes, exposure and outcome measures were self-reported.	No, baseline values, age, perceived social acceptance and intensity in social behaviour were included in the model.	Unclear how much loss to follow-up; formal attrition analyses revealed no differences between adolescents who did versus did not return for Wave 2 on any of the demographic or primary outcome measures in this study, with the exception of adolescents' ego development.	N/A
Tetzner, 2017	No, students were chosen using cluster sampling, which involved randomly selecting schools and then taking two full classes per school.	Yes, exposure and outcome measures were self-reported.	No, baseline values, peer acceptance, self-esteem and academic achievement were included in the model.	Unclear, no information on loss to follow-up available	N/A

Van Voorhees, 2008	No, each school that participated in the study was systematically chosen to represent urban, regional, and ethnic strata. After being stratified by grade and gender within each school, 17 adolescents were chosen randomly from each age–gender group (17/strata, 200/schools).	Yes, exposure and outcome measures were self-reported.	No, socio-demographic variables and baseline depression were included in the model.	Yes, 73.5% of the T1 participants completed the T2 survey, no attrition analysis reported.	N/A
Vanhalst, 2013	Yes, convenience sample of 3 schools; all children of grade 9 and 10 were invited. Lack of parental consent resulted in in exclusion of less than 1% of the potential sample	Yes, exposure and outcome measures were self-reported.	No, (baseline) self-esteem, social acceptance, loneliness, gender and grade were included in the model.	No, at each measurement wave, about 4% of the adolescents did not give their assent to participate in the study and were not included in the sample. Due to new students in the schools at each time point, a slightly increasing number of students participated in our study at each time point (so-called 'drop-in'). A statistical test indicated that missing data were completely ad random.	N/A
Wagner, 2018	Unclear, based on additional resources, a multi-sampling process seemed to be used to compose the initial larger samples, but no further details were found. Only data from one of the two	Yes, exposure and outcome measures were self-reported.	No, gender and grade were included as covariates.	Unclear, no information on loss to follow-up available	N/A

	German states was used for the current study. The subsamples were compared with the larger samples and it was shown that students from the subsample on average reported lower self-esteem and better grades (for grade 5 students) and slightly lower social self-concept scores and better grades (for grade 8 students).				
Webb, 2016	Yes, unclear how schools were selected ("first three consenting schools were permitted") and how participants were selected (probably all children of grades 5-7 were invited, but there was a low response rate).	Yes, exposure and outcome measures were self-reported.	No, (baseline) outcomes, peer acceptance, popularity, peer victimization and age were included in the model.	Unclear, no information on loss to follow-up available	N/A
Yen, 2015	No, all participants fulfilling the selection criteria and who had consent from their parents were included.	Yes, exposure and outcome measures were self-reported.	No, the model contained the following covariates: models contained covariates (black race, positive affect, and aggression).	No, At T2 83% of the T1 children were retained; there were no significant demographic or clinical differences between those who remained in the study and those who did not.	Adolescents and parents were compensated ($50) for the time it took to complete both the baseline interview and the follow-up interview.

Certainty of the body of evidence

<u>**5-11 years**</u>

Peer acceptance	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	

Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

12-18 years

Peer acceptance	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Social acceptance/social competence	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Peer connectedness	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Peer invalidation	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/large variability of the results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

5-18 years

Peer acceptance	Initial grading Low [C]	Downgrading due to

Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Social acceptance/social competence	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	**5-11 years** *Peer acceptance* There is limited evidence <u>with benefit</u> for peer acceptance. It was shown that peer acceptance/social inclusion resulted in a statistically significant increase in self-esteem (Wagner 2018), and a statistically significant decrease of depression, internalizing symptoms and externalizing symptoms (Klima 2008). However, in an additional study a statistically significant increase in self-esteem and depression could not be demonstrated (Kingery 2011). In addition, a statistically significant change in aggression (Hughes 2001), loneliness (Kingery 2007, Kingery 2011), school involvement (Kingery 2007, Kingery 2011) and self-worth (Klima 2008) could not be demonstrated. Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data. **12-18 years** *Peer acceptance* There is limited evidence <u>with benefit</u> for peer acceptance. It was shown that peer acceptance/social inclusion resulted in a statistically significant increase in self-esteem 1 to 3 school years later (Wagner 2018, Tetzner 2017) and a statistically significant decrease of depression (Van Voorhees 2008) and social withdrawal (Barzeva 2020). However, a statistically significant increase in self-esteem after 6 months (within 1 school year) could not be demonstrated (Tetzner 2017). In addition, the significant effect on social withdrawal was dependent of the measurement method (reported by the parent or the adolescent himself) (Barzeva 2020). Lastly, a statistically significant decrease in social anxiety could not be demonstrated (Barzeva 2020). Evidence is of very low certainty. *Social acceptance/social competence* There is limited evidence <u>with benefit</u> for social acceptance/social competence. It was shown that perceived social acceptance/social competence resulted in a statistically significant decrease in (social) anxiety (Teachman 2017, Grills-Taquechel 2010) and loneliness (Vanhalst 2013). Evidence is of very low certainty and results cannot be considered precise due to limited sample size. *Peer connectedness*

	There is limited evidence concerning the risk of marihuana use or other illegal drug use in case of peer connectedness. A statistically significant decreased risk of marihuana use or other illegal drug use in case of peer connectedness could not be demonstrated (McDonough 2016). Evidence is of very low certainty and results cannot be considered precise due to lack of data. *Peer invalidation* There is limited evidence <u>with harm</u> for peer invalidation. It was shown that peer invalidation resulted in a statistically significant increase in suicidal ideation (Selby 2013) and in self-harm in girls (Yen 2015). A statistically significant increase in suicide events in girls and boys, and in self-harm in boys could not be demonstrated (Yen 2015). Evidence is of very low certainty and results cannot be considered precise due to limited sample size or large variability of the results. **5-18 years** *Peer acceptance* There is limited evidence concerning body dysmorphic disorder symptoms following peer acceptance in children from 9 to 14 years old. A statistically significant decrease in body dysmorphic disorder symptoms in case of peer acceptance could not be demonstrated (Webb 2016). Evidence is of very low certainty and results cannot be considered precise due to lack of data. *Social acceptance/social competence* There is limited evidence <u>with benefit</u> for social acceptance/social competence in children between 8 and 13 years old. It was shown that perceived social acceptance/social competence resulted in a statistically significant decrease in aggression and depression (Brendgen 2004). Evidence is of very low certainty.
Reference(s)	**Articles** Barzeva SA, Richards JS, Meeus WHJ, Oldehinkel AJ. *The social withdrawal and social anxiety feedback loop and the role of peer victimization and acceptance in the pathways*. Dev Psychopathol 2020, 32(4):1402-1417 Brendgen M, Vitaro F, Turgeon L, Poulin F, Wanner B. *Is there a dark side of positive illusions? Overestimation of social competence and subsequent adjustment in aggressive and nonaggressive children*. J Abnorm Child Psychol 2004, 32(3):305-20 Grills-Taquechel AE, Norton P, Ollendick TH. *A longitudinal examination of factors predicting anxiety during the transition to middle school*. Anxiety Stress Coping 2010, 23(5):493–513 Hughes JN, Cavell TA, Prasad-Gaur A. *A positive view of peer acceptance in aggressive youth risk for future peer acceptance*. *Journal of School Psychology 2001, 39*(3):239–252 Kingery JN, Erdley CA. *Peer experiences as predictors of adjustment across the middle school transition*. *Education & Treatment of Children 2007, 30*(2):73–88 Kingery JN, Erdley CA, Marshall KC. Peer acceptance and friendship as predictors of early adolescents' adjustment across the middle school transition. *Merrill-Palmer Quarterly 2011, 57*(3):215–243 Klima T, Repetti RL. *Children's peer relations and their psychological adjustment: Differences between close friendships and the larger peer group*. *Merrill-Palmer Quarterly 2008, 54*(2):151–178 McDonough MH, Jose PE, Stuart J. *Bi-directional Effects of Peer Relationships and Adolescent Substance Use: A Longitudinal Study*. J Youth Adolesc 2016, 45(8):1652-63 Selby EA, Yen S, Spirito A. *Time Varying Prediction of Thoughts of Death and Suicidal Ideation in Adolescents: Weekly Ratings over Six Month Follow-Up*. J Clin Child Adolesc Psychol 2013, 42(4):481–495 Teachman BA, Allen JP. *Development of Social Anxiety: Social Interaction Predictors of Implicit and Explicit Fear of Negative Evaluation*. J Abnorm Child Psychol 2007, 35(1):63–78 Tetzner J, Becker M, Maaz K. *Development in multiple areas of life in adolescence: Interrelations between academic achievement, perceived peer acceptance, and self-esteem*. *International Journal of Behavioral Development 2017, 41*(6):704–713

Van Voorhees BW, Paunesku D, Kuwabara SA, Basu A, Gollan J, Hankin BL, Melkonian S, Reinecke M. *Protective and vulnerability factors predicting new-onset depressive episode in a representative of U.S. adolescents*. J Adolesc Health 2008, 42(6):605-16

Vanhalst J, Luyckx K, Scholte RHJ, Engels RCME, Goossens L. *Low self-esteem as a risk factor for loneliness in adolescence: perceived - but not actual - social acceptance as an underlying mechanism*. J Abnorm Child Psychol 2013, 41(7):1067-81

Wagner J, Lüdtke O, Robitzsch A, Göllner R,Trautwein U. *Self-esteem development in the school context: The roles of intrapersonal and interpersonal social predictors*. Journal of Personality 2018, 86:481–497

Webb HJ, Zimmer-Gembeck MJ, Mastro S. *Stress exposure and generation: A conjoint longitudinal model of body dysmorphic symptoms, peer acceptance, popularity, and victimization*. Body Image 2016, 18:14-8

Yen S, Kuehn K, Tezanos K, Weinstock L M, Solomon J, Spirito A. *Perceived family and peer invalidation as predictors of adolescent suicidal behaviors and self-mutilation*. *Journal of Child and Adolescent Psychopharmacology 2015, 25*(2): 124–130

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Peer acceptance	Barzeva, 2020	Classmate affection subscale and Classmate behavioral affirmation subscale of the Social Production Function	This self-reported measure includes items such as "Many classmates like to do things together with me", Many classmates help me if there is something I need help with," and "I can really trust my class-mates."
	Hughes, 2001	Subscale of the Pictorial Scale of Perceived Competence and Social Acceptance for Young Children	Questions: has lots of friends, others share their toys, has friends to play with, has friends on playground, gets asked to play with others, others sit next to you
(perceived) Social acceptance (among peers)	Brendgen, 2004 Grills-Taquechel, 2010 Teachman, 2007 Vanhalst, 2013	Social acceptance subscale of the Self-Perception Profile for Children/Adolescents	The subscale is comprised of six items, which were scored from 1 to 4 with higher scores reflecting a more positive self-image. This is a self-perceived measure. Example questions: "Some teens find it hard to make friends" Additional information: the "Social acceptance" subscale has been renamed to "Social competence" subscale in 2012, and now contains items referring to knowing how to make friends, having the skills to get others to like oneself, knowing what to do to have others like or accept you, understanding what it takes to become popular, etc. The previous social subscale, labeled "Social Acceptance" could be confounded with "social support" because it did not specify the role of the *self* in producing social outcomes. The previous version of this subscale is currently not available anymore.

Thematic category: Social rejection

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Agoston, 2013, USA	Observational: prospective cohort study (Study 1: 3 timepoints with 1- year intervals; Study 2: 3 timepoints with 6-month intervals)	Study 1 (middle childhood): 636 students from grades 2-4 from schools in several Midwestern towns, mean age 7.97±0.37 years Study 2 (early adulthood): 605 students from grades 5-8 from schools in several Midwestern towns, mean age 11.74±0.68 years	Peer rejection: teacher-rated on a 7-point scale (*Not At All* to *Extremely*): "How much is this child rejected by his/her peers?" Peer neglect: teacher-rated on a 7-point scale (*Not At All* to *Extremely*): "How much is this child neglected or overlooked by his/her peers?"	Outcomes measured: - depressive symptoms (self-reported) assessed by 'Short Mood and Feelings Questionnaire'. - aggression (teacher-rated) assessed by the 'Children's Social Behavior Scale' (CSBS). Statistics: path analysis was performed, using the full information maximum likelihood (FIML) estimation method to handle missing data. Each variable was included at each wave of assessment.
Christ, 2017, USA	Observational: prospective cohort study (3 timepoints with 18 month intervals, the 4th timepoint 4 years after the first timepoint)	2776 adolescents, aged 11 to 17 years (mean age 13.50±0.047 (SE) years) at any of the four waves of observation. All were in contact with the US Child Protective Services.	Peer isolation: latent variable with 7 adolescent self-reported questions from the Loneliness and Social Dissatisfaction Questionnaire for Young Children about relationships with peers	Outcomes measured: depressive symptoms assessed with Youth Self-Report questionnaire. Statistics: Structural equation modelling
Cotter, 2016, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	5894 adolescents in final model at T4 in 6th to 11th grade from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina, mean age 12.9 (boys) and 12.7 years (girls) at baseline (SD not available)	Peer rejection: self-reports as part of the School Success Profile Plus (SSP+), with 3 items, each 3 response options	Outcomes measured: - Internalizing symptoms (i.e., depression and anxiety) were measured with 7 items from the Youth Self Report - Externalizing behavior was measured using the modified 12-item externalizing subscale from the Youth Self Report Statistics: Hierarchical Linear Modeling analyses among male and female students were conducted separately on each of the 15 imputed data sets

Cotter, 2017, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	3580 adolescents in final model at T4 in 6th to 11th grade from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina, mean age was 12.04±1.53 years at baseline and 14.83±1.49 years at T4	Peer rejection: self-reports as part of the School Success Profile Plus (SSP+), with 3 items, each 3 response options.	Outcomes measured: aggression (self-reported) assessed using modified subscales from the Youth Self-Report (YSR). Statistics: 2-level hierarchical linear model
Demol, 2020, Belgium	Observational: prospective cohort study (3 timepoints with 3 t0 4-month intervals)	692 students, aged 7.92 - 13.14 years (mean age 10.28±0.94 years) in grades 4-6 from 10 primary schools	Peer rejection or the degree to which children are disliked by the peer group was measured by one peer nomination item: "Which classmates do you like the least?"	Outcomes measured: - relational peer victimization - physical peer victimization both assessed with two scales of the validated Social Experiences Questionnaire (SEQ) Statistics: Cross-lagged analyses to investigate the hypothesized longitudinal links
Di Giunta, 2018, Italy	Observational: prospective cohort study (5 timepoints with 1-year intervals)	607 adolescents students, aged 10 at baseline and reassessed at 11, 12, 13 and 14 years, from 2 public schools in Genzano, a residential community 30 km from Rome. Last follow-up at age 16-17 (mean age and SD unknown)	Peer rejection: At each time point (age 10 to 14 years old), children nominated the three most liked and the three most disliked classmates to play with, as well as the three most liked and the three most disliked classmates to study or do homework with.	Outcomes measured at age 16-17 : The Youth Self Report (YSR) questionnaire was used to measure the following subscales: - anxiety-depression - social competence - overt antisocial behaviour Statistics: path analysis using age 14 predictor data for age 16-17 outcomes
Evans, 2019, USA	Observational: prospective cohort study (3 timepoints with 1-year intervals)	912 children, aged 6 to 12 years (mean age and SD unknown), from an elementary school in the US Midwest. Data were collected annually from 7 consecutive grade cohorts of children who were assessed at up to 3 time points over the course of 3 school years (T1, T2 and T3).	Peer rejection: measured using teachers' ratings on 4 items from the Teacher Report Form (TRF)	Outcomes measured: depressive symptoms (teacher-rated), assessed using the withdrawn/depressed subscale from the Teacher Report Form (TRF). Statistics: path analysis using Rejection data in Grade 1 as predictor for Depressive symptoms in Grade 3 or Rejection data in Grade 3 as predictor for

				Depressive symptoms in Grade 5. This study follows an accelerated longitudinal design using transformed cross-sequential data and planned missingness.
Fussner, 2018, USA	Observational: prospective cohort study (2 timepoints with 6-month interval)	133 school children, aged 8 to 13 years (mean age 10.06±1.30 years), from 3rd – 6th grade of a Midwestern elementary school. Final sample consisted of 131 children.	Peer rejection: the Peer Relations Scale (PRS) was completed by teachers to assess children's peer rejection (including 9 items). All items are scored on a 4-point Likert scale; higher scores indicate greater peer rejection.	Outcomes measured: depressive symptoms (self-reported), assessed by the 25-item Revised Child Anxiety and Depression Scale—Short Version (RCADS-S) Statistics: Linear regression analysis
Kaltiala-Heino, 2010, Finland	Observational: prospective cohort study (2 timepoints with 2-year interval)	3278 students aged 15 years (mean age 15.5±0.4 years) at T1. The final sample consisted of 2070 adolescents who completed the surveys at both T1 and T2.	Peer isolation: respondent was asked how often the respondent had been left alone by peers against her/his wishes (self-reported). The response alternatives were: "many times a week or about once a week" vs "less frequently or not at all"	Outcomes measured: depression, assessed using R-BDI, a Finnish modification of the 13-item Beck Depression Inventory Statistics: Logistic regression analyses
Ladd, 2006, USA	Observational: prospective cohort study (7 timepoints in 1-year intervals, longitudinal study)	399 children, aged 5 to 12 years, recruited as they entered kindergarten (mean age 5.62 years, SD unknown) and followed until they completed 6th grade (mean age 11.71 years, SD unknown) in urban, suburban, and rural US locations	Peer rejection: consisting of three measures, used as indicators of this construct: 1) roster-and-rating measures (RR-SMS), asking ''How much do you like to play with this person in school?''; 2) peer nomination procedure (PN-NEG) in which children nominated up to three classmates by circling names on a roster; 3) teacher-rated Excluded by Peers subscale of the Child Behaviour Scale (CBS-EP)	Outcomes measured: - Externalizing problems, consisting of 3 measures: 1) classroom disruptiveness (CD) 2) Hyperactive-distractible subscale of the Child Behaviour Scale (CBS-HD); 3) Teacher Report Form Delinquent Behavior subscale (TRF-DB). - Internalizing problems, including measures of anxiety and depression, using the 17 items from the TRF Anxious/Depressed subscale (TRF-AD) and the 4 items from the Anxious Fearful subscale of the Child Behaviour Scale (CBS-AF). Statistics: prediction model with peer rejection as an additive risk factor

Miller-Johnson, 2002, USA	Observational: prospective cohort study (4 timepoints with 1-year intervals)	657 high risk children, from grade 1 to 4 (age range, mean age and SD unknown), from 4 schools in high-risk communities, selected based on indicators of poverty and crime	Peer rejection: assessed by peers' "like least" nominations (children were allowed to nominate an unlimited number of peers)	Outcomes measured: conduct problems were from parent ratings on the Diagnostic Interview Schedule for Children (DISC), used to diagnose conduct disorder (CD) and oppositional defiant disorder (ODD). Statistics: Structural equation modeling with maximum likelihood estimation was used to test the hypothesized mediation effects (rejection assessed in Grades 2 and 3, predicting conduct problems in Grade 4).
Pedersen, 2007, Canada	Observational: prospective cohort study (8 timepoints with 1-year intervals)	551 French Canadian children, mean age 6.10±0.30 to 13.09±0.30 years, from kindergarten to grade 7 in 5 elementary schools from a small community in northwestern Quebec, Canada	Peer rejection: assessed from ages 8 to 11 through peer-nomination by circling the names of the 3 children they liked most and 3 children they liked least. A composite score of peer rejection at ages 10 to 11 was used to predict outcomes at ages 12 to 13.	Outcomes measured: - depressive symptoms (self-reported) assessed from ages 12 to 13 by the Children's Depression Inventory (CDI) - loneliness (self-reported) assessed from ages 12 to 13 for participants' feelings of loneliness and social dissatisfaction - delinquency (self-reported) assessed from ages 12 to 13 by the Self-Reported Delinquency Questionnaire (SRDQ) Statistics: path analyses using the restricted model, in which paths that were not significant at the trend level or better were eliminated to improve model fit
Prinstein, 2004, USA	Observational: prospective cohort study (2 timepoints with 17-month interval)	158 adolescents, aged 15 to 17 years (mean age 16.31±0.50 years), from a suburban high school	Social preference: measured using a peer-nomination sociometric procedure. The difference between "like most" and "like least" standardized scores were restandardized, with higher scores indicating greater acceptance among peers,	Outcomes measured: depressive symptoms (self-reported) assessed at T1 and T2 using the Children's Depression Inventory

			and lower scores indicating greater peer rejection	Statistics: hierarchical linear regression analysis
Selby, 2013, USA	Observational: prospective cohort study (weekly ratings over a 6-month period)	119 adolescents after a psychiatric hospitalization for elevated suicide risk, aged 12 to 18 years (mean age 15.3±1.4 years)	Peer invalidation: following questions were used: "Were there times when you did not feel accepted by your classmates? Or that you were being left out? Or that you could not express your true thoughts or feelings? Or that if you did express your thoughts and feelings, that you would be dismissed, punished, ignored, or made fun of? How many friends do you have that can confide in?"	Identified from reference list from 'Yen 2015' Outcomes measured: suicidal ideation (self-reported) based on the following questions: "Sometimes when people are upset or feel bad they think about dying or even killing themselves. Do you have these thoughts during these past six months? How often did you have these thoughts? When you had them, how long did they last – a few minutes, an hour or more, almost always? Did you have these thoughts throughout the past month or where there times when you did not have these thoughts? Do you have those thoughts now? Do you have a plan? What is it? Have you told anyone about these thoughts or plans?" Statistics: hierarchical model with lag variables
Smokowski, 2016, USA	Observational: prospective cohort study (3 timepoints with 1-year intervals)	The final analytic sample comprised 4,065 observations at baseline, 4,251 observations at T 2, and 4,256 observations at T 3, from students 6th to 10th grade (mean age and SD not available) from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina.	Peer rejection: 3 self-reported items measured in year 1, 2 and 3 using the questions: "I am made fun of by my friends" and "I wish my friends would show me more respect." with the following response options: "Not like me; a little like me; a lot like me"	Outcomes measured: aggression, assessed with Youth Self-Report (YSR), the adolescent version of the Child Behaviour Checklist (CBC) Statistics: An imputation model with more than 70 variables was used to fill in the missing values. Subsequent hierarchical linear model (HLM) analyses were based on the 10 and 20 imputed files generated in the multiple imputation.

				Peer rejection was used as one of the time-varying covariates.
Smokowski, 2017, USA	Observational: prospective cohort study (4 timepoints with 1-year intervals)	The final analytic sample for internalizing symptoms comprised 3,751 observations at baseline, 3,981 observations at T 2, 4,839 observations at T 3 and 4,216 observations at T 4. The final sample used in the analysis for aggressive behavior was comprised of 3,735 observations at baseline, 3,999 observations at T 2, 4,872 observations at T 3, and 4,175 observations at T 4. Students 6th to 11th grade from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina. Mean age at baseline was 12.77±1.05 years.	Peer rejection: 3 self-reported items measured in year 1, 2 and 3 using the questions: "I am made fun of by my friends" and "I wish my friends would show me more respect." with the following response options: "Not like me; a little like me; a lot like me"	Outcomes measured: - aggression, assessed with 12-item subscale of Youth Self-Report (YSR) - internalizing symptoms, assessed with seven items from YSR that assess symptoms of anxiety and depression

Statistics: hierarchical linear model (HLM) analyses. Peer rejection was used as one of the time-varying covariates. |
| Snyder, 2012, USA | Observational: prospective cohort study (2 timepoints in 3- to 4-year intervals) | 267 students, mean age of 5.3 years at initial data collection and 9.3 years at last data collection from a community sample from one elementary school (SD unknown) | Peer rejection: assessed at last year of kindergarten (K) or Grade 1 using three indicators; 1) coercive peer behaviour, 2) degree of scapegoating, 3) accusation of wrongdoing by peers. The three indicators were standardized and then averaged to create the peer rejection construct. | Outcomes measured: Antisocial behaviour assessed in Grade 4 using Child Behavior Checklist, Teacher Report Form, and Youth Self-Report instruments

Statistics: path analysis using mediated model of skillful parenting, peer rejection, and their interaction separately for boys and girls |
| Vaske, 2010, USA | Observational: prospective cohort study (3 timepoints with 2-year intervals) | Children were aged 11 to 12.5 years (mean age 140.73 months, SD unknown) at T1 and a subsample that is referred to as the Children of | Peer rejection: assessed at age 13-14.5 years and consisted of two items that asked mothers to report how often their child 1) has trouble getting along with other children and 2) is not liked by other children. A | Outcomes measured: Delinquency at age 15-16.5 years (composite of 13 items)

Statistics: regression models were |

		the National Longitudinal Survey of Youth (NLSY) was selected for this study. Number of children unknown.	composite scale was created from the two items and this measure was dichotomized.	estimated separately for males and females

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
<td colspan="5" style="background:green">**5-11 years**</td>				
Internalizing problems (T1 --> T2)	Peer rejection	Not statistically significant: β: 0.11 £ (p>0.01)	1, 399 §	Ladd, 2006
Internalizing problems (T2 --> T3)		Statistically significant: β: 0.18 £ (p<0.01) *With harm for peer rejection*		
Internalizing problems (T3 --> T4)		Statistically significant: β: 0.37 £ (p<0.01) *With harm for peer rejection*		
Internalizing problems (T4 --> T5)		Statistically significant: β: 0.28 £ (p<0.01) *With harm for peer rejection*		
Internalizing problems (T5 --> T6)		Statistically significant: β: 0.23 £ (p<0.01) *With harm for peer rejection*		
Internalizing problems (T6 --> T7)		Statistically significant: β: 0.32 £ (p<0.01) *With harm for peer rejection*		
Externalizing problems (T1 --> T2)		Statistically significant: β: 0.27 £ (p<0.01) *With harm for peer rejection*		
Externalizing problems (T2 --> T3)		Statistically significant: β: 0.47 £ (p<0.01) *With harm for peer rejection*		
Externalizing problems (T3 --> T4)		Statistically significant: β: 0.36 £ (p<0.01) *With harm for peer rejection*		
Externalizing problems (T4 --> T5)		Statistically significant: β: 0.23 £ (p<0.01) *With harm for peer rejection*		
Externalizing problems (T5 --> T6)		Statistically significant: β: 0.18 £ (p<0.01) *With harm for peer rejection*		
Externalizing problems (T6 --> T7)		Statistically significant: β: 0.20 £ (p<0.01) *With harm for peer rejection*		
Relational victimization (T1 --> T2)		Statistically significant: β: 0.11 £ (p<0.01) *With harm for peer rejection*	1, 692	Demol, 2020

Relational victimization (T2 --> T3)		Not statistically significant: £† (p>0.05)		
Physical victimization (T1 --> T2)		Statistically significant: β: 0.11 £ (p<0.01) *With harm for peer rejection*		
Physical victimization (T2 --> T3)		Not statistically significant: £† (p>0.05)		
Depressive symptoms (T1$_{Grade1}$ --> T3$_{Grade3}$)		Not statistically significant: B: 0.214±0.104, 95%CI [-0.001;0.411] £† (p>0.05)	1, 155 §	Evans, 2019
Depressive symptoms (T1$_{Grade3}$ --> T3$_{Grade5}$)		Not statistically significant: B: 0.099±0.097, 95%CI [-0.093;0.286] £† (p>0.05)	1, 136 §	
Depressive symptoms (T1 --> T2)		Statistically significant: B: 2.29±0.65 £ (p<0.01) *With harm for peer rejection*	1, 131 §	Fussner, 2018
Depressive symptoms (T1 --> T2)		Not statistically significant: β: -0.13 £† (p>0.05)	1, 636	Agoston, 2013
Depressive symptoms (T2 --> T3)		Not statistically significant: β: 0.08 £† (p>0.05)		
Agression (T1 --> T2)		Not statistically significant: β: -0.01 £† (p>0.05)		
Agression (T2 --> T3)		Statistically significant: β: 0.13 £ (p<0.05) *With harm for peer rejection*		
Conduct problems (T2-3 --> T4)		Not statistically significant: β: -0.09±0.05 £† (p<0.10)	1, 657	Miller-Johnson, 2002
Antisocial behaviour (T$_{K-1}$ --> T$_{grade4}$)		*Boys:* Not statistically significant: β: 0.07 £† (p>0.10)	1, 134 §	Snyder, 2012
		Girls: Not statistically significant: β: 0.12 £† (p>0.10)	1, 133 §	
Depressive symptoms (T1 --> T2)	Peer neglect in middle childhood	Not statistically significant: β: 0.02 £† (p>0.05)	1, 636	Agoston, 2013
Depressive symptoms (T2 --> T3)		Not statistically significant: β: 0.04 £† (p>0.05)		
Agression (T1 --> T2)		Not statistically significant: β: -0.07 £† (p>0.05)		
Agression (T2 --> T3)		Not statistically significant: β: -0.11 £† (p>0.05)		
12-18 years				
Aggression	Peer rejection	Not statistically significant: B: 0.0050 (calculated from exp(B)) £† (p>0.05)	1, 3580	Cotter, 2017

		Statistically significant: B: 0.023 (calculated from exp(B)) £ (p<0.001) *With harm for peer rejection*	1, 4256	Smokowski, 2016
Internalizing symptoms		Statistically significant: B: 0.031 (calculated from exp(B)) £ (p<0.001) *With harm for friend rejection*	1, 4175	Smokowski, 2017
		Statistically significant: B: 0.049 (calculated from exp(B)) £ (p<0.001) *With harm for friend rejection*		
		Boys Statistically significant: B: 0.055 (calculated from exp(B)) £ (p<0.001) *With harm for peer rejection* *Girls:* Statistically significant: B: 0.046 (calculated from exp(B)) £ (p<0.001) *With harm for peer rejection*	1, 2894 1, 3000	Cotter, 2016
Externalizing symptoms		*Boys:* Statistically significant: B: 0.041 (calculated from exp(B)) £ (p<0.001) *With harm for peer rejection* *Girls* Statistically significant: B: 0.030 (calculated from exp(B)) £ (p<0.001) *With harm for peer rejection*	1, 2894 1, 3000	Cotter, 2016
Anxiety-depression ($T_{age14} \rightarrow T_{age16-17}$)		*Boys:* Statistically significant: β: 0.20 £ (p<0.01) *With harm for peer rejection* *Girls:* Not statistically significant: £† (p>0.05)	1, 274 § 1, 212 §	Di Giunta, 2018
Overt antisocial behaviour ($T_{age14} \rightarrow T_{age16-17}$)		*Boys:* Statistically significant: β: 0.18 £ (p<0.05) *With harm for peer rejection* *Girls:* Not statistically significant: £† (p>0.05)	1, 274 § 1, 212 §	Di Giunta, 2018
Delinquency (T2 --> T3)		*Boys:* Statistically significant: B: 0.481 £ (p≤0.05)	1, [numbers not reported] †	Vaske, 2010

		With harm for peer rejection		
		Girls: Not statistically significant: B: 0.169 £† (p>0.05)		
Depressive symptoms (T1 --> T2)	Peer isolation	*Boys* Not statistically significant: OR: 1.7, 95%CI [0.4;6.7] ¥ (p=0.47)	1, 903	Kaltiala-Heino, 2009
		Girls Not statistically significant: OR: 1.3, 95%CI [0.4;3.8] ¥ (p=0.65)	1, 1167	
	Social preference	Not statistically significant: β: -0.06 £† (p>0.05)	1, 137 §	Prinstein, 2004
Suicidal ideation (T$_{week\_n-1}$ --> T$_{week\_n}$)	Peer invalidation	Statistically significant: β: 0.11±0.03 £ (p<0.001) *With harm for peer invalidation*	1, 119 §	Selby, 2013
5-18 years				
Depressive symptoms (T1 --> T2)	Peer rejection	Not statistically significant: β: -0.01 £† (p>0.05)	1, 605	Agoston, 2013
Depressive symptoms (T2 --> T3)		Not statistically significant: β: -0.06 £† (p>0.05)		
Depressive symptoms (T$_{age10-11}$ --> T$_{age12-13}$)		Not statistically significant: £† (p>0.05)	1, 551	Pedersen, 2007
Agression (T1 --> T2)		Statistically significant: β: 0.17 £ (p<0.05) *With harm for peer rejection*	1, 605	Agoston, 2013
Agression (T2 --> T3)		Not statistically significant: β: 0.12 £† (p>0.05)		
Loneliness (T$_{age10-11}$ --> T$_{age12-13}$)		Statistically significant: β: 0.23 £ (p<0.05) *With harm for peer rejection*	1, 551	Pedersen, 2007
Delinquency (T$_{age10-11}$ --> T$_{age12-13}$)		Not statistically significant: £† (p>0.05)		
Depressive symptoms (T1 --> T2)	Peer neglect in early adolescence	Not statistically significant: β: 0.04 £† (p>0.05)	1, 605	Agoston, 2013
Depressive symptoms (T2 --> T3)		Statistically significant: β: 0.15 £ (p<0.05) *With harm for peer neglect*		
Agression (T1 --> T2)		Not statistically significant: β: -0.13 £† (p>0.05)		
Agression (T2 --> T3)		Not statistically significant: β: -0.08 £† (p>0.05)		
Depressive symptoms	Peer isolation	Statistically significant: B: 0.352±0.043 £ (p<0.001) *With harm for peer isolation*	1, 2776	Christ, 2017

B ± SE (unless otherwise indicated), OR: odds ratio, SD: standard deviation, SE: standard error, B: unstandardized coefficient, β: standardized coefficient

£ No SE and/or CI available; or no information on magnitude of effect available to assess variability of results
¥ Imprecision (large variability of results)
† Imprecision (lack of data)
§ Imprecision (limited sample size or low number of events)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Agoston, 2013	Unclear, families of 2nd graders of schools in several Midwestern towns were contacted, unclear if selection was random.	No, peer rejection, peer neglect and aggression were reported by teachers. However, depressive symptoms were self-reported.	No, path analyses accounted for baseline symptoms.	Yes, there was missing data up to 15% and students with complete data differed significantly for several variables from those with missing data. However, analyses were conducted using full information maximum likelihood estimation (FIML) to handle missing data.	No, no indication
Christ, 2017	Yes, data drawn from National Survey of Child and Adolescent Well-Being, no random sampling.	Yes, exposure and outcome self-reported.	No, all covariates were time-varying variables that were observed at each wave of the study, unless otherwise noted. Child characteristic covariates included sex, age, race/ ethnicity, and cognitive disability status.	No, missing data limited to 10% and observations with missing data on variables were included in the analyses in the SEM models and in the creation of factor scores using direct maximum likelihood with a conditional missing at random assumption.	No, no indication
Cotter, 2016	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40% of middle-school	Yes, the study relied exclusively on self-report measures and therefore represents adolescents' individual perceptions.	No, model was developed and adjusted for covariates	Unclear, missing data ranged from 0.19 to 0.35 from T1 to T4 and was not completely at random and all the analysis variables were used in multiple imputation to ensure the representativeness of covariance structure in the imputed data.	No, no indication

	students was included.				
Cotter, 2017	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40% of middle-school students was included.	Yes, the study relied exclusively on self-report measures and therefore represents adolescents' individual perceptions.	No, all variables except for demographics and indirect aggression (which was only measured at T4) were included as time varying covariates.	Unclear, in each wave, additional students entered the study.	Given the unique racial/ethnic makeup of the rural sample, results should be cautiously generalized to other settings.
Demol, 2020	Yes, no random sampling, children from 10 primary schools participated. No information on how schools were selected.	Yes, peer rejection was peer-nominated, outcome measures were self-reported.	No, age and gender effects were controlled for.	No, missing data limited to 10% and to deal with missing data, the models were analysed using the full information maximum likelihood (FIML) algorithm, which uses all available data instead of imputing scores.	No, no indication
Di Giunta, 2018	Yes, no random sampling, children recruited from only 2 schools.	Yes, peer rejection was peer-nominated, outcome measures are self-reported.	No, path analyses accounted for baseline variables and age.	Yes, follow-up assessment was very limited (only 60% of the original sample).	No, no indication
Evans, 2019	Yes, no random sampling, only students from one elementary school included.	No, exposure and outcomes were teacher-reported.	No, path analyses accounted for baseline depressive symptoms and gender.	No, Bayesian cross-classified estimation was used to account for planned-missing data.	- A cross-section of participants of different age cohorts is followed longitudinally to examine developmental trends across both time and age. Results do not apply to the same individuals over the entire period, as is required for true mediation. - Teachers received $7 per survey at T1 and $50 for full participation at T2/T3.

Fussner, 2018	Yes, no random sampling, children from only one elementary school included.	Yes, depressive symptoms were self-reported. Peer rejection was teacher-reported.	No, linear regression analyses accounted for baseline depressive symptoms and separate analyses for gender were conducted.	No, 2.3% missing data. There were no differences on any primary T1 variables or associations with demographic variables between those with and those without T2 data.	Teachers were compensated $5 for each packet completed; children received a grade-appropriate book at each time point.
Kaltiala-Heino, 2009	No, the Youth and Mental Health Study is a population-based, representative (cluster sampling), prospective cohort study.	Yes, exposures and outcomes were self-reported.	No, in the final model, depression at T1 is controlled for, as well as age, parental education (T1), and family structure (T1).	Yes, only 2070 of 3278 students completed both timepoints. Drop-outs differed from those who participated. Less drop-outs lived with parents or and had at least one parent with academic education. Drop-outs were somewhat older, and there were more boys than girls among them. Depression at T1 was more common in drop-outs. Those dropping out reported more often being bullies at T1 than those participating at the follow-up.	No, no indication
Ladd, 2006	Yes, no random sampling, children and their parents were recruited at kindergarten preregistration meetings that were held in school systems within multiple urban, suburban, and rural U.S. locations.	No, peer rejection was peer-nominated or teacher-reported.	No, the additive model included predictive values for aggression or withdrawn behaviour.	Unclear, 96.49% was followed-up until grade 6, no information on differences between drop-outs and those who completed all timepoints.	No, no indication
Miller-Johnson, 2002	Yes, no random sampling, schools from high-risk communities were selected based on indicators of poverty and crime.	No, peer rejection measured by peer nomination. Outcome measure was parent-rated.	No, data for aggression, ADHD and social preference were included in the model.	Yes, analyses indicated that sociometric data at Grade 2 or 3 was disproportionately missing for children who had lower social preference scores in first grade.	Yes, no transparency on measures of predictor and outcome data on other timepoints. Parents were compensated financially for their time, and

					children received a price.
Pedersen, 2007	Yes, no random sampling, children from only 5 schools in a small community were included.	Yes, peer rejection was assessed by peer nomination, the outcomes were self-reported.	No, demographic covariates sex and maternal occupational prestige, as well as early behavioural predictors were included in the model.	No, only 163 of the 551 children in the final sample had complete data across all variables created for the study. Between 70% and 84% of data were available for each study variable. However, there were no significant differences.	No, no indication
Prinstein, 2004	Yes, no random sampling, participants from the 10th grade of one high school were included.	Yes, peer rejection assessed by peer nomination, but outcome measures were self-reported.	No, the regression model controlled for T1 levels of depressive symptoms on an initial step, followed by gender and scores for attributional style (CASQ) and social preference on a second step	No, missing data (13%) at T2 were imputed with an expectation–maximization procedure, which utilized available self- and peer-reported data at T2, as well as all data available at T1. Data were missing completely at random according to Little's test.	No, no indication
Selby, 2013	Yes, no random sampling, participants were recruited from one psychiatric hospital.	Yes, exposure and outcome measures were self-reported.	No, baseline outcomes and several other factors were included in the model.	No, no loss to follow-up.	Adolescents and parents were compensated for their time with a payment of $50 to each, for the baseline interview and for the 6-month follow-up interview.
Smokowski, 2016	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40% of middle-school students was included.	Yes, exposure and outcome were self-reported.	No, model was developed and adjusted for covariates (demographics, positive proximal processes, and negative proximal processes (measured at baseline))	No, multiple imputation analyses were completed to minimize the impact of missing data, which could occur due to participant non-response to questions or due to attrition from the sample over time. An imputation model with more than 70 variables was used to fill in the missing values. Subsequent HLM analyses	Given the unique racial/ethnic makeup of the rural sample, results should be cautiously generalized to other settings.

				were based on the 10 and 20 imputed files generated in the multiple imputation.	
Smokowski, 2017	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40% of middle-school students was included.	Yes, exposure and outcome were self-reported.	No, model was developed and adjusted for covariates (15 predictors which can be categorized into the following three categories: (a) demographics, (b) microsystem protective factors, and (c) microsystem risk factors.	No, multiple imputation analyses were completed to minimize the impact of missing data.	Given the unique racial/ethnic makeup of the rural sample, results should be cautiously generalized to other settings.
Snyder, 2012	Yes, no random sampling, kindergarten children enrolling in one elementary school were recruited.	No, objectively measured exposure and outcomes.	No, several baseline variables were included in the model	Unclear, there was a substantial amount (17%) of missing data in Grade 4 but no information on differences between drop-outs and those who did participate.	No, no indication
Vaske, 2010	Yes, no random sampling, data used from the Children of the National Longitudinal Survey of Youth	Yes, peer rejection was reported by the mother, but outcome was self-reported.	No, all multivariate analyses controlled for child's age, race, and disobedience.	Unclear, no information on missing data available.	Yes, no transparency on the included population (e.g. demographic information)

Certainty of the body of evidence

5-11 years

	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

12-18 years

Peer rejection	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	

		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Peer isolation	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Large variability of the results
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Social preference	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Peer invalidation	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

5-18 years

Peer rejection	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	Lack of data, only in non-significant outcomes
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Peer neglect	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	Lack of data, only in non-significant outcomes
Inconsistency	0	
Indirectness	0	

Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Peer isolation	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Conclusion	**5-11 years** There is limited evidence with <u>harm</u> for **peer rejection**. It was shown that peer rejection resulted in a statistically significant increased risk of *externalizing symptoms* (Ladd 2006). It was shown that peer rejection at ages 6-11 resulted in a statistically significant increased risk of *internalizing symptoms* at ages 7-12. However, a statistically significant increased risk of internalizing symptoms at age 6, in case of peer rejection at age 5, could not be demonstrated (Ladd 2006). It was shown that peer rejection in 3rd grade resulted in a statistically significant increased risk of *aggression* in 4th grade. However, a statistically significant increased risk of aggression in 3rd grade, in case of peer rejection in 2nd grade, could not be demonstrated (Agoston 2013). In addition, it was shown that peer rejection in 4th grade resulted in a statistically significant increased risk of *relational or physical victimization* in 5th grade. However, a statistically significant increased risk of relational or physical victimization in 6th grade, in case of peer rejection in 5th grade, could not be demonstrated (Demol 2020). Furthermore, it was shown in one study that peer rejection resulted in a statistically significant increased risk of *depressive symptoms* (Fussner 2018). However, in 2 other studies, a statistically significant increased risk of depressive symptoms, in case of peer rejection, could not be demonstrated (Agoston 2013, Evans 2019). In addition, a statistically significant increased risk of *conduct problems* (Miller-Johnson 2002) or *antisocial behaviour* (Snyder 2012), in case of peer rejection, could not be demonstrated. Evidence is of very low certainty and results could not be considered precise due to limited sample sizes and/or lack of data. There is limited evidence concerning the risk of *depressive symptoms* or *aggression* in case of **peer neglect**. A statistically significant increased risk of depressive symptoms or aggression, in case of peer neglect, could not be demonstrated (Agoston 2013). Evidence is of very low certainty and results could not be considered precise due to lack of data. **12-18 years** There is limited evidence with <u>harm</u> for **peer/friend rejection**. It was shown that peer/friend rejection resulted in a statistically significant increased risk of *internalizing symptoms* (Cotter 2016, Smokowski 2017). In addition, it was shown that peer rejection resulted in a statistically significant increased risk of externalizing symptoms (Cotter 2016). Also, it was shown in 2 studies that peer rejection resulted in a statistically significant increased risk of *aggression* (Smokowski 2016, 2017). However, in one other study, a statistically significant increased risk of aggression, in case of peer rejection, could not be demonstrated (Cotter 2017).

| | Furthermore, it was shown that peer rejection resulted in a statistically significant increased risk of *anxiety-depression* and *overt antisocial behaviour* (Di Giunta 2018) or *delinquency* (Vaske 2010) in boys. However, in girls, a statistically significant increased risk of anxiety-depression, overt antisocial behaviour or delinquency, in case of peer rejection, could not be demonstrated (Di Giunta 2018, Vaske 2010). Evidence is over very low certainty and results cannot be considered precise due to limited sample sizes and/or lack of data.

There is limited evidence concerning the risk of depressive symptoms in case of **peer isolation** or **social preference.**
A statistically significant increased risk of *depressive symptoms*, in case of peer isolation or social preference, could not be demonstrated (Kaltiala-Heino 2009, Prinstein 2004).
Evidence is of very low certainty and results could not be considered precise due to

There is limited evidence with <u>harm</u> for **peer invalidation**.
It was shown that peer invalidation resulted in a statistically significant increased risk of *suicidal ideation* (Selby 2013).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size.

5-18 years
There is limited evidence with <u>harm</u> for **peer rejection**.
It was shown that peer rejection resulted in a statistically significant increased risk of *loneliness* (Pederson 2007).
It was shown that peer rejection in 5th or 6th grade resulted in a statistically significant increased risk of *aggression* in 6th or 7th grade. However, a statistically significant increased risk of aggression in 7th or 8th grade, in case of peer rejection in 6th or 7th grade, could not be demonstrated (Agoston 2013).
Furthermore, a statistically significant increased risk of *depressive symptoms*, in case of peer rejection in early adolescence, could not be demonstrated (Agoston 2013, Pedersen 2007).
Also, a statistically significant increased risk of *delinquency* at 12-13 years, in case of peer rejection at 10-11 years, could not be demonstrated (Pederson 2007).
Evidence is of very low certainty.

There is limited evidence with <u>harm</u> for **peer neglect**.
It was shown that peer neglect in 6th or 7th grade resulted in a statistically significant increased risk of *depressive symptoms* in 7th or 8th grade. However, a statistically significant increased risk of aggression in 6th or 7th grade, in case of peer rejection in 5th or 6th grade, could not be demonstrated (Agoston 2013).
Also, a statistically significant increased risk of *aggression*, in case of peer neglect, could not be demonstrated (Agoston 2013).
Evidence is of very low certainty and results cannot be considered precise due to lack of data.

There is limited evidence with <u>harm</u> for **peer isolation**.
It was shown that peer isolation resulted in a statistically significant increased risk of *depressive symptoms* (Christ 2017).
Evidence is of very low certainty. |
| **Reference(s)** | **Articles**
Agoston AM, Rudolph KD. *Pathways From Depressive Symptoms to Low Social Status.* J Abnorm Child Psychol 2013, 41(2):295-308.
Christ SL, Kwak YY, Lu T. *The joint impact of parental psychological neglect and peer isolation on adolescents' depression.* Child Abuse Negl 2017, 69:151-162.
Cotter KL, Smokowski PR. *An Investigation of Relational Risk and Promotive Factors Associated with Adolescent Female Aggression.* Child Psychiatry Hum Dev 2017, 48(5):754-767.
Cotter KL, Wu Q, Smokowski PR. *Longitudinal Risk and Protective Factors Associated with Internalizing and Externalizing Symptoms Among Male and Female Adolescents.* Child Psychiatry Hum Dev 2016, 47(3), 472-485.
Demol K, Leflot G, Verschueren K, Colpin H. *Revealing the Transactional Associations among Teacher-Child Relationships, Peer Rejection and Peer Victimization in Early Adolescence.* J Youth Adolesc 2020, 49(11):2311-2326.
Di Giunta L, Pastorelli C, Thartori E, Bombi AS, Baumgartner E, Fabes RA, Martin CL, Enders CK. *Trajectories of Italian Children's Peer Rejection: Associations with* |

Aggression, Prosocial Behavior, Physical Attractiveness, and Adolescent Adjustment. J Abnorm Child Psychol 2018, 46(5):1021-1035.

Evans SC, Fite PJ. *Dual Pathways from Reactive Aggression to Depressive Symptoms in Children: Further Examination of the Failure Model.* Journal of Abnormal Child Psychology 2019, 47:85-97.

Fussner LM, Luebbe AM, Mancini KJ, Becker SP. *Emotion dysregulation mediates the longitudinal relation between peer rejection and depression: Differential effects of gender and grade.* International Journal of Behavioral Development 2018, 42(2):155-166.

Kaltiala-Heino R, Fröjd S, Marttunen M. *Involvement in bullying and depression in a 2-year follow-up in middle adolescence.* European Child & Adolescent Psychiatry 2009, 19(1):45-55.

Ladd GW. *Peer rejection, aggressive or withdrawn behavior, and psychological maladjustment from ages 5 to 12: an examination of four predictive models.* Child Dev 2006, 77(4):822-846.

Miller-Johnson S, Coie JD, Maumary-Gremaud A, Bierman K, Conduct Problems Prevention Research Group. *Peer rejection and aggression and early starter models of conduct disorder.* J Abnorm Child Psychol 2002, 30(3):217-230.

Pedersen S, Vitaro F, Barker ED, Borge AIH. *The timing of middle-childhood peer rejection and friendship: linking early behavior to early-adolescent adjustment.* Child Dev 2007, 78(4):1037-1051.

Prinstein MJ, Aikins JW. *Cognitive moderators of the longitudinal association between peer rejection and adolescent depressive symptoms.* J Abnorm Child Psychol 2004, 32(2):147-158.

Selby EA, Yen S, Spirito A. *Time varying prediction of thoughts of death and suicidal ideation in adolescents: Weekly ratings over 6-month follow up.* J Clin Child Adolesc Psychol 2013, 42(4):481-495.

Smokowski PR, Guo S, Cotter KL, Evans CBR, Rose RA. *Multi-level risk factors and developmental assets associated with aggressive behavior in disadvantaged adolescents.* Aggress Behav 2016, 42(3):222-238.

Smokowski PR, Guo S, Evans CBR, Wu Q, Rose RA, Bacallao M, Cotter KL. *Risk and protective factors across multiple microsystems associated with internalizing symptoms and aggressive behavior in rural adolescents: Modeling longitudinal trajectories from the Rural Adaptation Project.* Am J Orthopsychiatry 2017, 87(1):94-108.

Snyder JJ, Schrepferman LP, Bullard L, McEachern AD, Patterson GR. *Covert antisocial behavior, peer deviancy training, parenting processes, and sex differences in the development of antisocial behavior during childhood.* Dev Psychopathol 2012, 24(3):1117-1138.

Vaske J, Gehring K. *Mechanisms Linking Depression to Delinquency for Males and Females.* Feminist Criminology 2010, 5(1):8-28.

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Peer rejection	Cotter, 2016 Cotter, 2017	Selected items from School Success Profile Plus (SSP+)	3 items, each with 3 response options ("Not like me", "A little bit like me", or "A lot like me"). Example items are: "I am made fun of by my friends" and "I wish my friends would show me more respect." See Bowen_2008 for more information on SSP.
	Evans, 2019	Selected items from the Teacher Report Form (TRF)	4 items from the Teacher Report Form (TRF). Ratings were provided on a three-point Likert scale (1 = not true, 2 = somewhat or sometimes true, 3=very or often true) and then averaged for analyses. The peer rejection subscale includes four items that tap general aspects of poor peer relations and social difficulties, including being teased, left out, or not liked. See Teacher's Report Form.pdf for full TRF.
	Fussner, 2018	Peer Relations Scale	The 9-item Peer Relations Scale (PRS) is adapted from the Dishion Social Acceptance Scale designed to assess peer rejection. The PRS has two sets of items. The first five items are scored on a 4-point Likert scale ranging from "never" (1) to "very often" (4). All items start with the stem: "How often is this student..." Sample endings include: "chosen late in the process of picking teams for sports or play activities" and "intentionally avoided by several other students as friends or playmates." The final four items are also on a 4-point scale from "none" (1) to "most" (4). These items ask teachers to indicate, "How many of the others students in the class would say that this student..." with a sample ending such as "is someone I really don't like." Mean scores across all nine items were used. Higher scores indicate greater peer rejection.
	Demol, 2020	[No formal name of instrument]	The degree to which children are disliked by the peer group was measured by one peer nomination item: "Which classmates do you like the least?" An unlimited number of nominations was allowed, but self-nomination was not allowed. Proportion scores were calculated by dividing the number of received nominations by the total number of possible nominations.
	Di Giunta, 2017	[No formal name of instrument]	At each time point, children nominated the three most liked and the three most disliked classmates to play with, as well as the three most liked and the three most disliked classmates to study or do homework with. Standardized social preference score was obtained for each child by subtracting the average of the two standardized dislike items from the average of the two standardized like items. Scores were then recoded such that high values indicate high levels of peer rejection.
	Pedersen, 2007	[No formal name of instrument]	Peer rejection was assessed from ages 8 to 11 through peer-nomination sociometric procedure as measure of social preference. Names of all children in a given class were

			handed out to the participants. Two research assistants ensured that all participants recognized the names of all classmates by reading them aloud in front of the class. The children were then asked to circle the names of the three children they liked most (positive nominations) and the three children they liked least (negative nominations). The number of years children were in the rejected category was computed for two 2-year intervals (i.e., ages 8 – 9 and 10 – 11). Values ranged from 0 (never rejected) to 2 (rejected at both time points).
Peer isolation	Christ, 2017	Items from Loneliness and Social Dissatisfaction Questionnaire for Young Children.	The original measure has 16 items assessing loneliness and social dissatisfaction from peer relationships. The seven chosen items included: "I have nobody to talk to at school.", "I have lots of friends at school.", "I feel alone at school.", "I don't have anyone to play with at school.", "I feel left out of things at school.", "There are no kids at school that I can go to when I need help.", and "I don't have any friends at school." Items were coded from 1 (=never) to 5 (=always). See Asher_1984_Childrens Loneliness and Social Dissatisfaction Scale.pdf for full scale.
Peer invalidation	Selby, 2013	[No formal name of instrument]	Self-reported weekly ratings, recalled with the help of time anchors during the 6-month follow-up interview. Following questions were used: "Were there times when you did not feel accepted by your classmates? Or that you were being left out? Or that you could not express your true thoughts or feelings? Or that if you did express your thoughts and feelings, that you would be dismissed, punished, ignored, or made fun of? How many friends do you have that you can confide in?"
Social preferece	Prinstein, 2004	[No formal name of instrument]	Peer acceptance/rejection was measured using a peer-nomination sociometric procedure. Using a roster of all grade-mates, adolescents nominated an unlimited number of peers whom they "liked to spend time with the most" and "liked to spend time with the least." The difference between "like most" and "like least" standardized scores was computed and restandardized as a measure of social preference, with higher scores indicating greater acceptance among peers, and lower scores indicating greater rejection.

Thematic category: Social influence

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Allen, 2006, USA	Observational: prospective cohort study (2	177 7th and 8th graders from a public middle	Susceptibility to peer influence at T1 (measured as the	Outcomes measured: - drug and alcohol use causing problems at

	timepoints in a 1-year period)	school in the South-eastern US (mean age at T1 13.36±0.66 years; at T2 14.26±0.78 years and N=154)	percentage of instances where the target adolescent and their closest friend disagreed about a choice in a hypothetical dilemma they were presented with, in which the target adolescent changed his or her initial answer)	T2 (4 items from the Self-Perception Profile for Adolescents) - depressive symptoms at T2 (Child Depression Inventory, CDI) Statistics: hierarchical linear regression analysis
Blodgett Salafia, 2010, USA	Observational: prospective cohort study (4 timepoints in a 3-year period)	85 adolescent girls from 5th grade (at T1) in a primary school in a Midwestern city (mean age at T1 10.59±0.52 years, range 10-12 years)	Peer pressure to be thin at T1 (Peer Pressure to be Thin Scale)	Identified from the systematic review of Webb 2013 Outcome measured: body dissatisfaction at T2 (Body Dissatisfaction subscale of the Eating Disorders Inventory, EDI) Statistics: structural equation modelling (SEM) [Data from timepoints 3 and 4 (7th and 8th grade) were not extracted because no direct links with the RF at T1 were reported]
Brown, 2004, USA	Observational: prospective cohort study (2 timepoints in a 5-year period)	1313 participants from 6th grade (at T1) in schools in Kentucky (mean age at T1 12.5 years)	Peer pressure resistance at T1, 6th grade (7-item scale, developed for the purpose of the study)	Outcome measured: marijuana use at T2, 10th grade (number of times having smoked marijuana) Statistics: hierarchical multivariable regression analysis
Cotter, 2016, USA	Observational: prospective cohort study (4 timepoints in a 4-year period)	Students from grades 6-8 at schools in rural, economically disadvantaged counties in North Carolina. At T1: 2090 boys (mean age 12.9 years) and 2181 girls (mean age 12.7 years)	Peer pressure at T1-T4 (5-item scale, developed for the purpose of the study)	Outcomes measured: - internalizing symptoms (i.e. depression and anxiety) at T1-T4 (7 items from the Youth Self-Report, YSR) - externalizing behaviour at T1-T4 (modified externalizing subscale from the YSR) Statistics: hierarchical linear regression modelling Part of the same study as Cotter 2017 and Smokowski 2016, 2017
Cotter, 2017	Observational: prospective cohort study (4 timepoints in a 4-year period)	2536 girls (at T1) from grades 6-8 at schools in rural, economically disadvantaged counties in North Carolina (mean	Peer pressure at T1-T4 (School Success Profile, SSP)	Outcome measured: aggression (modified subscale of the YSR) Statistics: hierarchical linear (mixed) regression modelling

		age at T1 12.04±1.53 years)		Part of the same study as Cotter 2016 and Smokowski 2016, 2017
Defoe, 2018, the Netherlands	Observational: prospective cohort study (2 timepoints in a 1-year period)	602 adolescents at T1 (mean age at T1 13.50±1.23).	Perceived peer pressure at T1 (two questions on stealing and vandalism from the Peer Pressure Inventory, PPI)	Outcome measured: delinquency at T2 (vandalism and property crime; 7 items, among which 5 questions from the International Self-Reported Delinquency Questionnaire, ISRD) Statistics: multiple path analysis
Eamon, 2001, USA	Observational: prospective cohort study (2 timepoints in a 2-year period)	963 children at T1 (10-12 years old at T1)	Peer pressure at T1 (a dichotomous variable, indicating whether at least one pressure from a list of 5 possible pressures was experienced, e.g. try marijuana or skip school)	Identified from the reference list of Eamon 2004 Outcome measured: antisocial behaviour at T2 (4 items from the Behavior Problems Index, BPI) Statistics: hierarchical multiple regression
Eamon, 2004, USA	Observational: prospective cohort study (2 timepoints in a 2-year period)	807 children at T1 (10-12 years old at T1)	Peer pressure at T1 (at least one pressure from a list of 5 possible pressures, e.g. try marijuana or skip school)	Outcome measured: disruptive school behaviour at T2 (4 items, of which 2 from the BPI) Statistics: hierarchical multiple regression
McCabe, 2005, Australia	Observational: prospective cohort study (3 timepoints in a 16-month period)	494 boys (mean age 13.08 years) and 359 girls (mean age 12.89) from grade 7 at T1	Perceived pressure to lose weight from best male friend and best female friend at T1 (Sociocultural Influences on Body Change Questionnaire)	Identified from the systematic review of Webb 2013 Outcomes measured: - body dissatisfaction at T2 and T3 (Body Image and Body Change Questionnaire, BIBCQ) - extreme weight loss at T2 and T3 (BIBCQ) Statistics: repeated measures analysis of variance
McDonough, 2016, New Zealand	Observational: prospective cohort study (3 timepoints with 1-year intervals)	1940 students from 78 schools on the North Island of New Zealand who participated in at least 2 time points (mean age at T1 12.20±1.75 years, range 10-15 years)	Negative peer influence at T1 and T2 (3 questions on doing something wrong to please friends)	Outcomes measured: - marijuana use at T2 and T3 (Washington Healthy Youth Survey, self-reported) - other illegal drug use at T2 and T3 (Washington Healthy Youth Survey, self-reported) Statistics: cross-lagged panel models
Power, 2005, USA	Observational: prospective cohort study (4 timepoints with	1253 adolescents in the 9th to 12th grade from six	Peer involvement in prosocial activities (protective factor; 4-item Friends Models for	Outcome measured: problematic drinking (Quantity and frequency of consumption were

	6-month intervals)	high schools in a large metropolitan school district in the Southwest	Conventional Behavior scale)	tapped by questions on the frequency of drinking in the past year and month, frequency of intoxication, typical quantity of consumption and frequency of consumption of large quantities) Statistics: logistic regression models
Sijtsema, 2014, the Netherlands	Observational: prospective cohort study (2 timepoints with 1-year interval)	369 students from 2 age groups: 133 elementary school children (mean age at T1: 9.36±0.52 years) and 236 middle school children (mean age at T1: 11.91±0.73 years) from public schools in Northern Italy	- Direct friendship influence (total similarity effect, measured by combining friendship network (up to 5 peer nominations as "best friends") and these friends' perceived bullying behaviour at T1) - Indirect friendship influence (measured by combining friendship network and these friends' scores on a 14-item moral disengagement scale)	Outcome measured: perceived bullying behaviour displayed by up to 5 peers (Bully Scale from the Participant Role Questionnaire, PRQ) Statistics: longitudinal social network analysis
Smokowski, 2016, USA	Observational: prospective cohort study (3 timepoints with 1-year intervals)	Students from grades 6-8 at schools in rural, economically disadvantaged counties in North Carolina. At T1: 4065 participants (mean age 12.80±0.02 years)	Peer pressure at T1-T2 (5-item subscale of the SSP)	Outcome measured: aggression at T2 (modified subscale of the YSR) Statistics: hierarchical linear regression modelling Part of the same study as Cotter 2016, 2017 and Smokowski 2017
Smokowski, 2017, USA	Observational: prospective cohort study (4 timepoints with 1-year intervals)	Students from grades 6-8 at schools in rural, economically disadvantaged counties in North Carolina. At T1: 3981 participants (mean age 12.77±1.05 years)	Peer pressure at T1-T2 (5-item subscale of the SSP)	Outcomes measured: - internalizing symptoms (i.e. depression and anxiety) at T1-T4 (7 items from the YSR) - aggression at T1-T4 (modified subscale of the YSR) Statistics: hierarchical linear regression modelling Part of the same study as Cotter 2016, 2017 and Smokowski 2016

Snyder, 2008, USA	Observational: prospective cohort study (2 timepoints with 4-year interval)	267 children attending kindergarten of an elementary school serving a low-socioeconomic neighbourhood; mean age at T1 5.3 years, mean age at T2 8.9 (3th grade) or 9.3 years (4th grade	Peer coercion at T1, defined by 3 variables: 1) rate at which peers directed coercive verbal and physical behaviour toward the target child; 2) scapegoating by peers (Likert-scale ratings completed by Family and Peer Process Code (FPPC) coders); 3) accused of wrongdoing by peers (FPPC coders)	Identified from the reference list of Snyder 2012 (evidence summary on social rejection) Outcomes measured: - overt conduct problems at T2 (aggression scales from parent Child Behavior Checklist, CBCL; Teacher Report Form, TRF; YSR) - covert conduct problems at T2 (delinquency scales from parent CBCL; TRF; YSR) Statistics: SEM
Tsakpinoglou, 2017, Canada	Observational: prospective cohort study (2 timepoints with a 1-year interval)	294 students (mean age at T1 15.38 years) from 8 French-speaking schools in Québec, Canada	Perceived friend pressure at T1 (6 items from the PPI)	Outcome measured: marijuana use at T2 ("How many times have you used marijuana or hashish in the past month?") Statistics: hierarchical multiple regression
Weymouth, 2018, USA	Observational: prospective cohort study (3 timepoints with 1-year intervals)	416 adolescents from 6th grade in schools in the southeast (mean age at T1 11.86±0.69 years, range 11-14 years)	Compliance to peers at T2 (11 items from the Resistance to Peer Influence Scale)	Outcomes measured: - social anxiety at T3 (Social Anxiety Scale for Children-Revised, SASC-R) - depressive symptoms at T3 (CDI) Statistics: SEM
Whitesell, 2014, USA	Observational: prospective cohort study (4 timepoints with 6-month intervals)	381 students (age range 11 to 14+ years) from the 6th and 7th grade at middle schools in an American Indian reservation in the Northern Plains, one of the poorest regions of the US	- Deviant peer influence - Prosocial peer influence	Outcome measured: marijuana use in the past month Statistics: multinomial logistic regression models

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
5-12 years				
Friend and peer influence				
Bullying behaviour T1 (9 y) → T2 (10 y)	Total similarity (direct influence)	Not statistically significant: B: 1.227±1.012 £† (p>0.05)	1, 133 §	Sijtsema, 2014
	Friends' moral disengagement (indirect influence)	Not statistically significant: B: 0.038±0.213 £† (p>0.05)		
Peer coercion				

Overt conduct problems T1 (5 y) → T2 (9 y)	Observed peer coercion	Statistically significant: B: 0.24 £ (p<0.05) *With harm for peer coercion*	1, 267 §	Snyder, 2008
Covert conduct problems T1 (5 y) → T2 (9 y)		Not statistically significant: B: 0.07 £† (p>0.05)		

Friend and peer pressure

Marijuana use T1 (15 y) → T2 (16 y)	Friend pressure	Not statistically significant: aOR: 0.86 £† (p>0.05)	1, 294 §	Tsakpinoglou, 2017
Delinquency T1 (13.5 y) → T2 (14.5 y)	Peer pressure	Not statistically significant: B: 0.01±0.01, 95%CI [-0.017;0.029] £† (p=0.49)	1, 602	Defoe, 2018
Externalizing behaviour (aggression) T1 (12 y) → T3 (14 y)		Statistically significant: B: 0.023±0.001 (p<0.001) *With harm for peer pressure*	1, 4065	Smokowski, 2016
Externalizing behaviour (aggression) T1 (12 y) → T4 (15 y)		Girls: Statistically significant: B: 0.039 £ (p<0.001) *With harm for peer pressure*	1, 2536	Cotter, 2017
		Boys: Not statistically significant: B: 0.004 £† (p>0.05)	1, 2090	Cotter, 2016
		Girls: Not statistically significant: B: 0.002 £† (p>0.05)	1, 2181	
		Not statistically significant: B: 0.008 £† (p>0.05)	1, 3735	Smokowski, 2017
Internalizing symptoms T1 (12 y) → T4 (15 y)		Boys: Statistically significant: B: 0.029 £ (p<0.01) *With harm for peer pressure*	1, 2090	Cotter, 2016
		Girls: Not statistically significant: B: 0.023 £† (p>0.05)	1, 2181	
		Statistically significant: B: 0.031 £ (p<0.001) *With harm for peer pressure*	1, 3715	Smokowski, 2017
Extreme weight loss T1 (13 y) → T2 (13.7 y)	Pressure from male friend to lose weight	Girls: Statistically significant: β: 0.14 £ (p<0.05) *With harm for pressure from male friend to lose weight*	1, 359 §	McCabe, 2005
	Pressure from female friend to lose weight	Girls: Not statistically significant: β: 0.05 £† (p>0.05)		
	Pressure from male friend to increase muscles	Girls: Not statistically significant: β: -0.02 £† (p>0.05)		

	Pressure from female friend to increase muscles	Girls: Not statistically significant: β: 0.10 £† (p>0.05)	
Extreme weight loss T1 (13 y) → T3 (14.3 y)	Pressure from male friend to lose weight	Boys: Not statistically significant: β: -0.08 £† (p>0.05)	1, 494
		Girls: Not statistically significant: β: -0.07 £† (p<0.05)	1, 359 §
	Pressure from female friend to lose weight	Boys: Not statistically significant: β: -0.06 £† (p>0.05)	1, 494
		Girls: <u>Statistically significant:</u> β: 0.16 £ (p<0.05) *With harm for pressure from female friend to lose weight*	1, 359 §
	Pressure from male friend to increase muscles	Boys: Not statistically significant: β: 0.11 £† (p>0.05)	1, 494
		Girls: Not statistically significant: β: -0.11 £† (p>0.05)	1, 359 §
	Pressure from female friend to increase muscles	Boys: Not statistically significant: β: 0.05 £† (p>0.05)	1, 494
		Girls: <u>Statistically significant:</u> β: 0.13 £ (p<0.05) *With harm for pressure from female friend to increase muscles*	1, 359 §
Body dissatisfaction T1 (13 y) → T3 (14.3 y)	Pressure from male friend to lose weight Pressure from female friend to lose weight Pressure from male friend to increase muscles Pressure from female friend to increase muscles	Boys: Not statistically significant: F(10,293): 0.40 ££† (p>0.05)	1, 494
	Pressure from male friend to lose weight Pressure from female friend to lose weight Pressure from male friend to increase muscles	Girls: Not statistically significant: F(10,227): 1.41 ££† (p>0.05)	1, 359 §

	Pressure from female friend to increase muscles			
Peer pressure resistance				
Marijuana use T1 (12.5 y) → T2 (16.5 y)	Peer pressure resistance	Statistically significant: B: -0.63±0.06 (p<0.01) *With benefit for peer pressure resistance*	1, 1313	Brown, 2004
Friend and peer influence				
High risk drinking T1 → T4 (14-18 y, 18 months later)	Peer prosocial behaviour	Not statistically significant: OR: 1.40 £† (0.05<p<0.06)	1, 743	Power, 2005
Starting using marijuana	Prosocial peer influence	Not statistically significant: OR: 1.11 £† (p>0.05)	1, 246 §	Whitesell, 2014
Increasing marijuana use		Not statistically significant: OR: 0.53 £† (p>0.05)		
Starting using marijuana	Deviant peer influence	Statistically significant: OR: 2.32 £ (p<0.01) *With harm for deviant peer influence*		
Increasing marijuana use		Not statistically significant: OR: 1.75 £† (p>0.05)		
Bullying behaviour T1 (12 y) → T2 (13 y)	Friends' moral disengagement (indirect friend influence)	Not statistically significant: B: -0.439±0.292 £† (p>0.05)	1, 236 §	Sijtsema, 2014
	Total similarity (direct friend influence)	Not statistically significant: B: 2.299±1.455 £†		
Susceptibility to friend and peer influence				
Drug and alcohol use causing problems T1 (13 y) → T2 (14 y)	Susceptibility to peer influence	Statistically significant: B: 0.24 £ (p<0.01) *With harm for susceptibility to peer influence*	1, 177 §	Allen, 2006
Depression T1 (13 y) → T2 (14 y)		Statistically significant: B: 0.14 £ (p<0.05) *With harm for susceptibility to peer influence*		
Depressive symptoms T2 (12-15 y) → T3 (13-16 y)	Compliance to peers	Not statistically significant: β: 0.08±0.13 £† (p<0.05)	1, 329 §	Weymouth, 2018
Social anxiety T2 (12-15 y) → T3 (13-16 y)		Statistically significant: β: 0.24±0.11 (p<0.05) *With harm for compliance to peers*		
5-18 years				
Peer pressure				
Body image dissatisfaction T1 (10-12 y) → T2 (11-13 y)	Peer pressure to be thin	Statistically significant: B: 0.64 £ (p<0.05) *With harm for peer pressure to be thin*	1, 85 §	Blodgett Salafia, 2010
Antisocial behaviour T1 (10-12 y) → T2 (12-14 y)	Peer pressure	Statistically significant: B: 0.152±0.118 (p<0.001) *With harm for peer pressure*	1, 963	Eamon, 2001

Disruptive school behaviour T1 (10-12 y) → T2 (12-14 y)		Statistically significant: β: 0.13 £ (p<0.001) *With harm for peer pressure*	1, 807	Eamon, 2004

Friend and peer influence

Marijuana use T1 (10-15 y) → T2 (11-16 y)	Negative peer influence	Statistically significant: β: 0.46 £ (p<0.01) *With harm for negative peer influence*	1, 1940	McDonough, 2016
Marijuana use T2 (11-16 y) → T3 (12-17 y)		Statistically significant: β: 0.19 £ (p<0.01) *With harm for negative peer influence*		
Other illegal drug use T1 (10-15 y) → T2 (11-16 y)		Statistically significant: β: 0.48 £ (p<0.01) *With harm for negative peer influence*		
Other illegal drug use T2 (11-16 y) → T3 (12-17 y)		Statistically significant: β: 0.30 £ (p<0.01) *With harm for negative peer influence*		

B ± SE (unless otherwise indicated), aOR: adjusted odds ratio, SE: standard error, CI: confidence interval, B: unstandardized coefficient, β: standardized coefficient, F: test statistic of ANCOVA
£ No SE or CI available; or no information on magnitude of effect available to assess variability of results
££ No raw data available
† Imprecision (lack of data)
§ Imprecision (limited sample size)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Allen, 2006	Yes, one specific public middle school was involved; all students of 7th and 8th grade were invited and included (parental consent).	Yes, outcomes were self-reported.	No, gender, ethnic minority status, and baseline outcome values were controlled for in the multivariable regression model.	Unclear, 154 out of 177 participants at T1 were included at T2 (87%) but the characteristics of drop-outs were not compared to those who remained in the study.	Yes, susceptibility to peer influence was measured as susceptibility to influence of only one (best) friend, instead of a larger social network.
Blodgett Salafia, 2010	Yes, all parents of children in the 4th grade of primary schools in a certain city were contacted and self-selected to participate. Proportion of participants compared to	Yes, exposures and outcomes were self-reported.	No, maternal influence and BMI were inserted in the SEM.	No, due to attrition as well as missing data, data for 85 out of 102 dyads were available at the end of the study (85%), but these did not differ significantly on any demographic variables from	Yes, study sample was not representative in terms of ethnicity and SES (mainly White and upper-middle-class) Covariates such as pubertal status and media

	total population is unknown.			the dyads that dropped out.	influence were not included.
Brown, 2004	No, according to the study that describes the study population (Clayton 1996, all primary schools in a certain county were included in the study and 99% of the students were granted parental consent).	Yes, exposures and outcomes were self-reported.	No, sex, race, friends' use and baseline outcome value were controlled for in the multivariable regression model.	Yes, attrition was 45% after 5 years and drop-outs were much more likely to having used marijuana at baseline than those who remained in the study (Clayton 1996).	Yes, the important (according to the authors) covariate SES was not included in the model.
Cotter, 2016	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40 % of middle-school students was included.	Yes, the study relied exclusively on self-report measures and therefore represents adolescents' individual perceptions.	No, model was developed and adjusted for covariates.	Unclear, missing data ranged from 0.19 to 0.35 from T1 to T4 and was not completely at random and all the analysis variables were used in multiple imputation to ensure the representativeness of covariance structure in the imputed data.	Yes, given the unique racial/ethnic makeup of the rural sample, results should be cautiously generalized to other settings.
Cotter, 2017	No, see Cotter 2016	Yes, see Cotter 2016	No, all variables except for demographics and indirect aggression (which was only measured at T4) were included as time varying covariates.	Unclear, in each wave, additional students entered the study.	see Cotter 2016
Defoe, 2018	Unclear, 8 schools involved but not clear how these were selected; all students in grade 1 or 3 of secondary school were invited. 74% of the eligible students participated (parental consent).	Yes, exposures and outcomes were self-reported.	No, baseline outcome value, peer norms and quality of relationship with parents were included as covariates.	Unclear, according to the study that describes the study population (Defoe 2016), 15% dropped out after T1, but it is unclear whether drop-outs were different in the variables relevant for data interpretation.	Yes, father and sibling factors, and positive influences, were not included.
Eamon, 2001	Unclear, the original	Yes, exposures were self-	No, many relevant	Unclear, cases with missing	Yes, apart from being self-

	National Longitudinal Survey of Youth probably used random sampling, but it is unclear how the current sample of 1184 children (including those with missing data) was obtained.	reported and outcomes were parent-reported.	demographic, familial and social covariates were accounted for in the regression model.	data were deleted (221 out of 1184, 19%), but it is not reported whether they differ from cases that remained in the study.	reported, validity of the measures used has not been established.
Eamon, 2004	Unclear, see Eamon 2001	Yes, exposures and outcomes were self-reported.	No, see Eamon 2001	Unclear, cases with missing data were deleted (377 out of 1184, 32%), but it is not reported whether they differ from cases that remained in the study.	N/A
McCabe, 2005	No, schools were drawn from a wide range of socio-economic areas and equally divided between state and private schools (stratified random sampling); all children from grade 7 were invited. 56% of the eligible children participated (active parental consent).	Yes, exposures and outcomes were self-reported.	No, message variables were accounted for (from mother, father, media), and separate analyses were conducted for boys and girls.	No, 263 out of 853 participants dropped out, but there were no significant differences in outcomes between those who dropped out and those completed the questionnaire at each timepoint.	N/A
McDonough, 2016	Unclear, 78 schools involved but not clear how these were selected; unclear how participants from the schools were recruited.	Yes, exposure and outcomes were self-reported.	No, baseline outcomes and peer connectedness and peer influence were taken into account in the cross-lagged panel model.	No, data from youth who participated in at least two time points were used in the current study (10.8 % attrition from T1). The final sample was nationally representative for gender and SES.	Yes, according to the authors, other potential moderators should be explored. The statistical model contains some unexplained variance.
Power, 2005	No, randomly selected participants received letters	Yes, exposure and outcome measures were self-reported.	No, logistic regression models were used to adjust	Yes, overall, from the first to the fourth time point,	Yes, only significant relationships are shown in the

	and parental consent forms. However, it is unclear how the schools were selected.		for other risk and protective factors; control variables were gender, ethnicity and grade in school	approximately 60% of the students remained in the study. Dropouts had higher scores on sensation seeking, mother's attitudes toward adolescent alcohol use, peer involvement in antisocial behavior and hours worked and lower scores on self-esteem, peer security, peer involvement in prosocial activities and school commitment. Additionally, dropouts were less likely to be abstainers and more likely to be date or heavy, multiple context drinkers.	summarizing table (reporting bias).
Sijtsema, 2014	Yes, 3 schools involved but not clear how these were selected; participation of classrooms within schools was authorized by school principals and teacher committees.	Yes, exposure and outcome measures were self-reported.	No, models were controlled for structural network effects, sex selection effects, and baseline values of outcome.	Unclear, few missing data and no drop-out from T1 to T2, but it is not clear whether this is caused by perfect compliance or by an unreported removal of drop-outs from the T1 data.	N/A
Smokowski, 2016	No, see Cotter 2016	Yes, see Cotter 2016	No, model was developed and adjusted for covariates (demographics, positive proximal processes, and negative proximal processes (measured at baseline))	No, multiple imputation analyses were completed to minimize the impact of missing data, which could occur due to participant non-response to questions or due to attrition from the sample over time. An imputation model with more than 70 variables was	Yes, see Cotter 2016

				used to fill in the missing values. Subsequent HLM analyses were based on the 10 and 20 imputed files generated in the multiple imputation.	
Smokowski, 2017	No, see Cotter 2016	Yes, see Cotter 2016	No, model was developed and adjusted for covariates (15 predictors which can be categorized into the following three categories: (a) demographics, (b) microsystem protective factors, and (c) microsystem risk factors.	No, multiple imputation analyses were completed to minimize the impact of missing data.	Yes, see Cotter 2016
Snyder, 2008	Yes, one specific school was involved; all children of kindergarten were invited and included (parental consent).	Yes, peer coercion was measured by professional coders with only modest to good agreement, and conduct problems were partly scored by parent's ratings.	No, baseline values of outcomes and demographic characteristics were controlled for.	No, full information maximum likelihood estimation was used to minimize the impact of missing data.	N/A
Tsakpinoglou, 2017	Unclear, 8 schools involved but not clear how these were selected; it is not clear how children were recruited within schools either.	Yes, exposure and outcome measures were self-reported.	No, baseline values of outcomes, gender and social characteristics were controlled for.	Unclear, a subsample with complete data was used, which differed from the total population in family composition but not in gender. However, results for other variables were not reported.	Yes, homogeneous study sample (90% Caucasia, middle class). Measure of friendship influence processes relied solely on the perception of one member of the friendship dyad.
Weymouth, 2018	No, the study sample was reasonably representative of the county in terms of race, parents' marital status	Yes, exposure and outcome measures were self-reported.	No, several covariates were included in the SEM: pubertal status, race, negative life events, parent depressive	Unclear, number or % of drop-outs was not reported.	Yes, only two-parent families who were primarily European American were included, which limited generalizability.

	and family poverty status.		symptoms, family income, parent education.		
Whitesell, 2014	No, 7 schools were randomly selected from all 14 middle schools across the reservation.	Yes, exposure and outcome measures were self-reported.	No, logistic regression models were adjusted for gender, age, cultural engagement, early puberty and parent-child relationship.	Yes, number of participants dropped from 381 at T1 to 246 at T4 (35% drop-out), but characteristics of drop-outs and compliant participants were not compared.	Yes, study was performed within a single cultural context, and results should not be seen as indicative of patterns in other Native American communities, let alone in the general society.

Certainty of the body of evidence

5-12 years

Friend and peer influence	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Peer coercion	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

12-18 years

Friend and peer pressure	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Peer pressure resistance	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	

Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Friend and peer influence	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Susceptibility to friend and peer influence	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

5-18 years

Peer pressure	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Friend and peer influence	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	**5-11 years** *Friend and peer influence*

There is limited evidence concerning the risk of **bullying behaviour** in case of friends' moral disengagement (**indirect friendship influence**) or total similarity to friends (**direct friendship influence**).
A statistically significant increased risk of bullying behavior in case of direct or indirect friendship influence could not be demonstrated (Sijtsema, 2014).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data.

Peer coercion

There is limited evidence <u>with harm</u> for peer coercion.
It was shown that **coercion by peers** resulted in a statistically significant increase in **overt conduct problems** 4 years later (Snyder 2008). However, a statistically significant increase in **covert conduct problems** could not be demonstrated (Snyder 2008).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data.

12-18 years

Peer pressure

There is limited evidence <u>with harm</u> for peer pressure.
It was shown that **peer pressure** resulted in a statistically significant increase in **externalizing behaviour problems (aggression)** 2 or 3 years later (Smokowski 2016, Cotter 2017). However, similar analyses of almost the same study population could not demonstrate a statistically significant increase in aggression (Cotter 2016, Smokowski 2017).
Similarly, it was shown that peer pressure resulted in a statistically significant increase **in internalizing behaviour problems** 3 years later in boys and girls combined (Smokowski 2017) and in boys (Cotter 2016). The latter study could not demonstrate a statistically significant increase in internalizing behaviour problems in girls (Cotter 2016).
Finally, a statistically significant increased risk of **delinquency** in case of direct or indirect friendship influence could not be demonstrated (Sijtsema, 2014).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data.

There is limited evidence <u>with harm</u> for **pressure from friends to lose weight or increase muscles** in **girls**.
It was shown that perceived pressure from male friends to lose weight, resulted in a statistically significant increase in **extreme weight loss** in girls 8 months later (McCabe 2005). It was also shown that perceived pressure from female friends to lose weight or to increase muscles, resulted in a statistically significant increase in extreme weight loss in girls 16 months later (McCabe 2005). However, a statistically significant increase in **body dissatisfaction** 8 months or 16 months later in case of peer pressure to lose weight or to increase muscles, could not be demonstrated (McCabe 2005).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data.

There is limited evidence concerning the risk of **extreme weight loss** or **body dissatisfaction** in case of **pressure from friends to lose weight or increase muscles** in **boys**.
A statistically significant increase in extreme weight loss or body dissatisfaction 8 months or 16 months later in case of peer pressure to lose weight or to increase muscles, could not be demonstrated (McCabe 2005).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data.

Peer pressure resistance

There is limited evidence <u>with benefit</u> for **resistance against peer pressure**.
It was shown that peer pressure resistance resulted in a statistically significant decrease in **marijuana use** 4 years later (Brown 2004).

	Evidence is of very low certainty. *Friend and peer influence* There is limited evidence concerning the risk of **high risk drinking** or **marijuana use** in case of **prosocial peer influence**. A statistically significant decrease in high risk drinking or marijuana use in case of prosocial peer behaviour could not be demonstrated (Power 2005, Whitesell 2014). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. There is limited evidence with <u>harm</u> for deviant peer influence. It was shown that **deviant peer influence** resulted in a statistically significant increase in **starting using marijuana** (Whitesell 2014). However, a statistically significant **further increase in marijuana use** in case of deviant peer behaviour could not be demonstrated (Whitesell 2014). Furthermore, an increase in **bullying behaviour** in case of friends' moral disengagement (**indirect friendship influence**) or total similarity to friends (**direct friendship influence**) (Sijtsema 2014) could not be demonstrated. Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. *Susceptibility to friend and peer influence* There is limited evidence <u>with harm</u> for susceptibility to friend and peer influence. It was shown that **susceptibility to peer influence** resulted in a statistically significant increase in **problematic drug and alcohol use** and in **depression** (Allen 2006). Additionally, it was shown that **compliance to peers** resulted in a statistically significant increase in social **anxiety** (Weymouth 2018). However, a statistically significant increase in **depressive symptoms** in case of compliance to peers (Weymoth 2018) could not be demonstrated. Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. **5-18 years** *Peer pressure* There is limited evidence <u>with harm</u> for peer pressure. It was shown that **perceived peer pressure** resulted in a statistically significant increase in **antisocial behaviour** and **disruptive school behaviour** (Eamon 2001, 2004). It was also shown that **perceived peer pressure to be thin** resulted in a statistically significant increase in **body image dissatisfaction** (Blodget Salafia 2010) Evidence is of very low certainty. *Friend and peer influence* There is limited evidence <u>with harm</u> for friend and peer influence. It was shown that **negative peer influence** resulted in a statistically significant increase in **marijuana use** and **other illegal drug use** (McDonough 2016). Evidence is of very low certainty.
Reference(s)	**Articles** Allen JP, Porter MR, McFarland FC. *Leaders and followers in adolescent close friendships: susceptibility to peer influence as a predictor of risky behavior, friendship instability, and depression.* Dev Psychopathol 2006, 18(1):155-172. Blodgett Salafia EH, Gondoli DM. *A 4-Year Longitudinal Investigation of the Processes by Which Parents and Peers Influence the Development of Early Adolescent Girls' Bulimic Symptoms.* The Journal of Early Adolescence 2011, 31(3):390-414. Brown TL, Miller JD, Clayton RR. *The Generalizability of Substance Use Predictors Across Racial Groups.* The Journal of Early Adolescence 2004, 24(3):274-302. Cotter KL, Smokowski PR. *An Investigation of Relational Risk and Promotive Factors Associated with Adolescent Female Aggression.* Child Psychiatry Hum Dev 2017, 48(5):754-767. Cotter KL, Wu Q, Smokowski PR. *Longitudinal Risk and Protective Factors Associated with Internalizing and Externalizing Symptoms Among Male and Female Adolescents.* Child Psychiatry Hum Dev 2016, 47(3):472-485.

Defoe IN, Dubas JS, van Aken MAG. *The Relative Roles of Peer and Parent Predictors in Minor Adolescent Delinquency: Exploring Gender and Adolescent Phase Differences.* Front Public Health 2018, 6:242.
Eamon MK. *Poverty, Parenting, Peer, and Neighborhood Influences on Young Adolescent Antisocial Behavior*. Journal of Social Service Re-search 2001, 28(1):1-23.
Eamon MK, Altshuler SJ. *Can We Predict Disruptive School Behavior?* Children & Schools 2004, 26(1):23-37.
McCabe MP, Ricciardelli LA. *A prospective study of pressures from parents, peers, and the media on extreme weight change behaviors among adolescent boys and girls*. Behav Res Ther 2005, 43(5):653-668.
McDonough MH, Jose PE, Stuart J. *Bi-directional Effects of Peer Relationships and Adolescent Substance Use: A Longitudinal Study*. J Youth Adolesc 2016, 45(8):1652-1663.
Power TG, Stewart CD, Hughes SO, Arbona C. *Predicting patterns of adolescent alcohol use: a longitudinal study*. J Stud Alcohol 2005, 66(1):74-81.
Sijtsema JJ, Rambaran JA, Caravita SC, Gini G. *Friendship selection and influence in bullying and defending: effects of moral disengagement*. Dev Psychol 2014, 50(8):2093-2104.
Smokowski PR, Guo S, Cotter KL, Evans CB, Rose RA. *Multi-level risk factors and developmental assets associated with aggressive behavior in disadvantaged adolescents*. Aggress Behav 2016, 42(3):222-238.
Smokowski PR, Guo S, Evans CB, Wu Q, Rose RA, Bacallao M, Cotter KL. *Risk and protective factors across multiple microsystems associated with internalizing symptoms and aggressive behavior in rural adolescents: Modeling longitudinal trajectories from the Rural Adaptation Project.* Am J Orthopsychiatry 2017, 87(1):94-108.
Snyder J, Schrepferman L, McEachern A, Barner S, Johnson K, Provines J. *Peer deviancy training and peer coercion: dual processes associated with early-onset conduct problems.* Child Dev 2008, 79(2):252-268.
Tsakpinoglou F, Poulin F. *Best friends' interactions and substance use: The role of friend pressure and unsupervised co-deviancy.* J Adolesc 2017, 60:74-82.
Weymouth BB, Buehler C. *Early adolescents' relationships with parents, teachers, and peers and increases in social anxiety symptoms.* J Fam Psychol 2018, 32(4):496-506.
Whitesell NR, Asdigian NL, Kaufman CE, Big Crow C, Shangreau C, Keane EM, Mousseau AC, Mitchell CM. *Trajectories of substance use among young American Indian adolescents: patterns and predictors.* J Youth Adolesc 2014, 43(3):437-453.

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Peer coercion	Snyder, 2008	Family and Peer Process Code (FPPC)	3.5 h videotaped interactions. Videotapes were coded using the FPPC on 3 variables: 1) rate at which peers directed coercive verbal and physical behaviour toward the target child during); 2) scapegoating by peers (Likert-scale ratings); 3) accused of wrongdoing by peers. All ratings were completed by FPPC coders. Average coder agreement was 76% to 82% for the first variable, 72% for the second, and 76% for the third variable.
Friend pressure	Tsakpinoglou, 2017	Peer Pressure Inventory (PPI)	First, participants were asked to name their best friend. Second, 2 weeks later, they filled in the PPI with the best friend's name automatically inserted. The questionnaire consisted of 6 items on a 5-point Likert scale on friend pressure (e.g. "Sometimes I do certain things so that Paul won't think I'm afraid"). See Table 1 in the paper for all items.
Peer pressure	Defoe, 2018		Two selected items concerning stealing and vandalism: - peer pressure to not shoplift or steal anything vs to steal something

			- peer pressure to not trash things or vandalize property vs to trash or vandalize things. Scores ranged from -3 ("a lot of pressure not to steal/vandalize" over 0 ("no pressure") to +3 ("a lot of pressure to steal/vandalize").
	Cotter, 2016, 2017, Smokowski, 2016, 2017	School Success Profile (SSP)	5 items of the SSP on peer pressure, on a 3-point Likert scale ("not like me" – "a little like me" – "a lot like me") (e.g. "I let my friends talk me into doing things I really don't want to do")
	Eamon, 2001, 2004	(no formal instrument)	Children reported whether they felt pressure (yes/no) from friends in 5 items: try cigarettes, try marijuana/other drugs, drink alcohol, skip school, commit a crime/engage in violence. A dichotomous variable was constructed for peer pressure (1= at least one pressure; 0 = no pressure).
Peer pressure to lose weight	McCabe, 2005	Sociocultural Influences on Body Change Questionnaire	Subscales on Body Change Influences from Best Male Friend and Best Female Friend, each comprising 7 items on perceived pressures to lose weight (3 items) or increase muscles (4 items).
Peer pressure to increase muscles			
Peer pressure to be thin	Blodgett Salafia, 2010	Peer Pressure to be Thin Scale	Developed for this study; items were drawn from the McKnight Risk Factor Survey. Six items, scored from 0 ("never") to 5 ("always"), e.g. "How important has it been to your friends that you be thin?", "How often have you changed your eating when you were around your friends?"
Peer pressure resistance	Brown, 2004	(no formal instrument)	7-item scale, developed for the purpose of the study. E.g. "If your best friend was skipping school, would you skip too?", "If a friend asks you to smoke marijuana with them, would you do it?" Responses were made using a 5-point scale from 1 ("definitely not") to 5 ("definitely would").
Peer prosocial behaviour	Power, 2005	Friends Models for Conventional Behavior scale	4 items (e.g. school clubs and youth groups)
Deviant peer influence	Whitesell, 2014	(no formal instrument)	The number of friends who encouraged disobeying parents, encouraged dangerous behaviour, got in trouble at school, and got into a lot if fights. For each item, scores ranged from 1 ("none") to 5 ("almost all or all").
Prosocial peer influence			The number of friends who volunteered or participated in community groups, went to Inipi (sweat ceremony) or church regularly, thought schoolwork was very important, and planned to go to college. For each item, scores ranged from 1 ("none") to 5 ("almost all or all").
Negative peer influence	McDonough, 2016	(no formal instrument)	Youth were asked how often they had "gone against the wishes of adults (e.g., parents, teachers) to make your friends happy," "done badly at something (e.g., schoolwork, sport) just to please your friends," and "done something that could get you in trouble because your friends wanted you to do it." Response options ranged from 1 (never/almost never) to 5 (always/almost always).
Friend's moral disengagement	Sijtsema, 2014	14-item self-report measure designed for school children by Caprara et	Participants nominated up to 5 peers whom they considered as their "best friend" (within the study sample). For each participant, friends' moral disengagement was assessed by averaging scores of these friends. Participants rated their moral disengagement from detrimental conduct on a 5-

	al. (1995) (in Italian)		point scale, from 1 ("strongly disagree") to 5 ("strongly agree"). E.g. "It is okay to treat somebody badly who behaved like a disgusting being".
Compliance to peers	Weymouth, 2018	Resistance to Peer Influence Scale	Adapted version of the scale: 11 items (e.g. "I say things I don't really believe because I think my friends will respect me more") and a 4-point response format (from 1="not like me" to 4="a lot like me"). The questionnaire was completed by the adolescent, father, mother and teacher and these manifest scores were used to calculate the latent measure of compliance to peers.

Thematic category: Social support

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Arora, 2017, USA	Observational: prospective cohort study (2 waves with a 1-year interval)	186 5th-through 8th-grade Asian American students (mean age 12.50±1.16 years) in a Midwest area comprised of over 100 urban, suburban, and rural schools	Support from peers (5-item questionnaire) at T1	Outcomes measured at T1 and T2: - Depressive symptoms: adapted measure based on the 20-item Center for Epidemiologic Studies Depression Scale (CES-D) Statistics: Path analysis model
Auerbach, 2011, Canada	Observational: prospective cohort study (5 waves with a 6-week interval)	258 adolescents, aged 12-18 years (mean age 14.48±1.47 years), attending high schools in Montreal, Canada	Peer and classmate social support (The Social Support Scale for Children and Adolescents (SSSCA))	Outcomes measured: - Depressive symptoms: The 20-item CES-D - Anxious symptoms: the 10-item Multidimensional Anxiety Scale for Children—Short Form (MASC-SF) - General mental health: Dependent interpersonal stress: The 57-item Adolescent Life Events Questionnaire-Revised Statistics: Multilevel modeling analyses
Branje, 2004, The Netherlands	Observational: prospective cohort study (3 waves, with 1-year interval)	288 families consisting of 2 parents and 2 of their biological children. Children were 11-15 years at T1 (mean age not available).	Perceived sibling support: measured with Relational Support Inventory (RSI).	Outcomes measured: - Problem behaviour, assessed with Nijmegen Problem Behavior List (NPBL) to measure internalizing and

		285 families still participated at T3.		externalizing behaviour
				Statistics: Step-wise hierarchical regression analysis
Branstetter, 2011, USA	Observational: prospective cohort study (2 waves with a 1-year interval)	166 adolescents in the 10th grade at baseline, aged 14-16 years (mean age 15.3 years, SD not available), attending public schools in both urban and suburban areas of a large metropolitan city in the Western United States	Friendship support (15 items of modified version of the Network of Relationships Inventory (NRI)) at T1	Outcomes measured with Drug Involvement Scale for Adolescents (DISA; self-report): - Marijuana use - Hard drug use (cocaine, opiates, depressants, tranquilizers, hallucinogens, inhalants, stimulants, over-the-counter drugs, and club drugs) Statistics: Hierarchical regression analyses
Burke, 2017, Switzerland	Observational: prospective cohort study (4 waves with a 6-month interval)	960 adolescents in the 7th grade, aged 11.1-15.3 years (mean age 13.2±0.59 years) at baseline, attending schools in 3 of the 26 Swiss Cantons	Friendship support at T1-T4 (selected items from the Inventory of Parent and Peer Attachment)	Outcomes measured at T1-T4: - Depressive symptoms: Depressive mood scale consisting of the following items in the analysis: "Sometimes I think everything is so hopeless that I do not feel like doing anything," "I think my life is kind of sad," and "I think my life is not worth living." Statistics: Trivariate cross-lagged model (TCLM)
Burton, 2004, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	496 adolescent girls, aged 11-15 years (mean age 13.0±0.73 years) at baseline, attending public and private middle schools in a metropolitan area of the southwestern United States	Peer support at T2 (12 items from the Network of Relationships Inventory (NRI))	Outcomes measured at T2-T4: - Depressive symptoms: The Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS). Responses were used to classify participants as having met threshold or subthreshold diagnostic criteria for major depression during the past year.

				Statistics: Hierarchical multiple regression models
Cavanaugh, 2015, USA	Observational: prospective cohort study (3 waves with a 1-year interval)	416 adolescents in the 6th grade at baseline, aged 11-14 (mean age 11.86±0.69 years), from 13 middle schools in a large, geographically diverse county in the south-eastern US	Peer support at T2 (5-item friend support scale of Richman 1997 and 7-item children's friendship scale of Berndt 1986)	Outcomes measured at T1 and T3: - Social anxiety symptoms: fears of negative evaluation (3-item subscale) and social avoidance of new peers (3-item subscale) using the Social Anxiety Scale for Children–Revised Questionnaire - Loneliness: Two parcels were created by randomly assigning 4 items to each parcel from the 8-item, UCLA Loneliness Scale. Loneliness also was measured using a single item, "I feel lonely" Statistics: Structural equation modeling
Colarossi, 2003, USA	Observational: prospective cohort study (2 waves with a 1-year interval)	217 adolescents, aged 15-18 (mean age 17 years, SD not available), from three school districts in suburban, midwestern communities	Peer support at T1 (Iowa Youth and Families Inventory).	Outcomes measured: - Depressive symptoms: 9-item self-report Symptoms Checklist-Revised (SLC-90-R) - General mental health: Self-esteem: 7-item self-report scale derived from Harter (1982) Statistics: Structural equation models
Cotter, 2016, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	5894 adolescents in final model at T4 in 6th to 11th grade, mean age 12.9 (boys) and 12.7 years (girls) at baseline (SD not available), from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina	Friend support (5-item scale from the School Success Profile questionnaire)	Outcomes measured: - Internalizing symptoms (i.e., depression and anxiety) were measured with 7 items from the Youth Self Report - Externalizing behavior was measured using the modified 12-item externalizing subscale from the Youth Self Report Statistics: Hierarchical Linear Modeling analyses among male and female students were conducted separately

				on each of the 15 imputed data sets
Cotter, 2017, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	3580 adolescents in final model at T4 in 6th to 11th grade, mean age was 12.04±1.53 years at baseline and 14.83±1.49 years at T4, from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina	Friend support: assessed with the School Success Profile	Outcomes measured: aggression, assessed with Youth Self-Report, the adolescent version of the Child Behaviour Checklist Statistics: Two-level hierarchical linear model (HLM)
Dauber, 2011, USA	Observational: prospective cohort study (2 waves with a 1-year interval)	2225 adolescent girls (1625 white and 600 African American) in the 7th- to 12th grade (mean age 15.4±0.11 years) were sampled within schools (sampling design of Add Health).	Peer support (research-made questionnaire) at T1	Outcomes measured: - Problematic drinking (self-report) Increasers included adolescents who transitioned from a low-risk class into one of the problematic drinking classes (moderate, heavy, or problem) Decliners: vice versa Statistics: Following latent class and transition analysis multinomial logistic regressions were performed.
Fanti, 2012, Cyprus	Observational: prospective cohort study (2 waves with a 1-year interval)	1416 adolescents, aged 11-14 years at T1 (mean age 12.89±0.78 years) from 13 middle schools	Supportive social relations (Friend support): measured with Multidimensional Scale of Perceived Social Support	Outcomes measured: - cyberbullying, assessed with The Student Survey of Bullying Behaviour-Revised Statistics: Hierarchical linear regression analysis
Gagné, 2020, Canada	Observational: prospective cohort study (3 waves over a 1-year period)	339 grade 7 (mean age 12.05 years, SD not available) and grade 10 (mean age 15.14 years, SD not available) students at baseline. Final sample of 240 students at T3	Perceived friends' support: measured using four items of the Multidimensional Scale of Perceived Social Support	Outcomes measured: appearance esteem, assessed with the Body-Esteem Scale. Statistics: Longitudinal multilevel growth modeling
Jaycox, 2009, USA	Observational: prospective cohort study (2 waves with a 6-month interval). Selection was based on the principle of a case-control study.	368 teens with or without depression, aged 13-18 years (mean age 15.17±0.17 (SE)), were recruited from different sites (free, public, private managed care, and private insurance medical offices and	Peer support (Social adjustment Scale for Youth) at T1	Outcomes measured: - Severity of depressive symptoms: Patient Health Questionnaire for Adolescents (sum of 10 depressive symptoms with "nearly every day" vs "for a few days" vs not experienced)

		clinics) in Los Angeles, California, and Washington, DC, metropolitan areas between Jan 2005 and March 2006		Statistics: Cross-lagged regression analyses by using multivariate regression
Jiang, 2012, USA	Observational: prospective cohort study (2 waves: fall 2008 and spring 2009)	614 grade 7 (mean age 12.68±0.67 years) and grade 8 (mean age 13.08±0.77 years) students from a large middle school in the Southeastern USA	Peer Support for Learning: assessed using the Peer Support for Learning subscale of the Student Engagement Inventory (SEI)	Outcomes measured: School satisfaction, assessed with the School Satisfaction subscale of the Multidimensional Students' Life Satisfaction Scale (MSLSS) Statistics: Two sets of hierarchical regression models
Grills-Taquechel, 2010, USA	Observational: prospective cohort study (2 waves with a 2-year interval)	77 youth in the 6th (mean age 11.69±0.52 years) and 8th grade (mean age 13.64±0.54 years) in a southwestern Virginia middle school	Support from classmates and close friends (Social Support Scale for Children, SSC) at T1	Outcomes measured at T2: - Anxiety: Multidimensional Anxiety Scale for Children, MASC (39-item self-report measure). Results are provided for four subscales: (1) Physical Symptoms (tense/restlessness and somatic/ autonomic); (2) Harm Avoidance (anxious coping and perfectionism); (3) Social Anxiety (humiliation/rejection and performing in public fears); (4) Separation Anxiety Statistics: Multiple hierarchical regression analyses
Kendrick, 2012, Canada/ Sweden	Observational: prospective cohort study (2 waves, approximately 1-year interval)	880 adolescents, aged 12-16 years (mean age 13.72±0.78) at T1, attending seven junior high schools (grades seven through nine) in a town in central Sweden	Supportive friends: Adolescents were instructed that they should nominate peers of any gender or age whom they considered as being very important in their lives, and with whom they talk, spend time, and do things. The youths were then asked to answer questions concerning support, trust, and behaviors in their relationships with their first-mentioned, or most important, peers. The items were	Outcomes measured: - bullying: participants were asked 3 questions: 'Have you said nasty things, mocked or teased anyone in an unpleasant way at school?'; 'Have you beaten, kicked, or assaulted anyone in an unpleasant way at school or on the way to or from school?'; and 'Have you participated in ostracizing someone?'. Each item was rated on a 4-point scale, with

			adapted from the validation and caring subscale of the Friendship Quality Questionnaire.	possible responses being: (1) 'No, it has not happened'; (2) 'Yes, it has happened once or twice'; (3) 'Yes, it has happened about once a week'; and (4) 'Yes, it has happened several times a week'. - Depression: scale was adapted from the Center for Epidemiologic Studies Depression Scale (CES-D), which is a 20-item self-report depression scale. For the present study, 16 of the original 20 items were used. Statistics used: Structural equation modeling with manifest variables.
Khatib, 2013, USA	Observational: prospective cohort study (2 waves with a 2-year interval)	821 pupils in the 7th (11-12 years) and 9th grade (13-14 years) at baseline (mean age was not available) in secondary schools in East London. There were 248 White British pupils, 344 Bangladeshi pupils and 229 Black pupils in the analyses.	Social support from friends (Multidimensional Scale of Perceived Social Support, MSPSS)	Outcomes measured: - General mental health: Psychological distress: self-report of Strengths and Difficulties Questionnaire (SDQ) - Depressive symptoms: self-report of Short Moods and Feelings Questionnaire (SMFQ) SDQ scores were dichotomized by scores equal to or above 17.5; and SMFQ values by a score equal to or above 8 at follow up. Statistics: Multivariate logistic regression analysis
Kirsch, 2007, Canada	Observational: case-control study	Cases: 25 patients (mean age 13.36 ± 0.76 years) with anorexia nervosa of the restricting subtype (n=22), anorexia nervosa of the bulimic subtype (n=1) or eating disorder not otherwise specified (n=2).	Conditional support by peers: assessed with a scare designed by Marold	Outcome: eating disorder, diagnosed at the emergency department using the Diagnostic Interview for Children and Adolescents-Revised (DICA-R) Statistics: Multivariate analyses of variance (MANOVA)

		Controls: age-matched and categorized according to their negative eating attitude scores on the Children's Eat Attitude Test (ChEAT). The low-risk sample (ChEAT<20, n=25, mean age 12.56±0.71 years) and high-risk (ChEAT≥20, n=25, mean age 12.42±0.82 years)		
Klima, 2008, USA	Observational: prospective cohort study (2 waves with a 2-year interval)	247 children (primarily Caucasian population) in the 4th grade (mean age 9.5 years, SD not available) at baseline attending parochial and public schools in a large metropolitan area	Close friend support (Close Friend Support subscale of the Social Support Scale for Children)	Outcomes measured: - General mental health: Global self-worth: The 6-item Global Self-Worth subscale of the Self-Perception Profile for Children - General mental health: Internalizing and externalizing symptoms: The 36 and 34-item Teacher Report Form's (TRF) internalizing and externalizing broadband scale, respectively. - Depressive symptoms: the 27-item Children's Depression Inventory (CDI) (self-report) Statistics: Multiple hierarchical regression analyses
Lester, 2012, Australia	Observational: prospective cohort study (4 waves between 2005 and 2007: grade 7, beginning and end of grade 8 and end of grade 9)	3549 students from 21 secondary schools (grade 7 (mean age 11 years, SD not available), beginning of grade 8, end of grade 8 (mean age 13 years, SD not available) and end of grade 9 (mean age 14 years, SD not available)	Peer support: assessed with peer support at school scale (adapted from the 24-item Perceptions of Peer Social Support Scale; Ladd et al., 1996)	Outcome measures: Bullying perpetration-victimization: - Bullying perpetration was assessed using a 9-item category index derived from items used in Rigby & Slee (1998), Olweus (1996) and the 2004 Youth Internet Survey. A perpetration score at each time-point was calculated for each student by averaging the 9 perpetration items (higher score indicating more

| | | | | perpetration experiences).
- Victimisation was assessed using a similar 9-item victimisation index which asked students how often they were bullied by others in the ways listed to measure perpetration.
A perpetration–victimisation score at each time-point was calculated for each student by averaging the perpetration and victimisation items, (higher score reflecting more overall bullying experiences). Only students who reported both perpetrating bullying and being victimised at least once or twice in the previous term are included in the analyses.

Statistics used: Cross-lagged models within the Structural Equation Modeling (SEM) framework were used to model causal paths, between factors of interest and perpetration–victimisation with longitudinal data collected over and following the students' transition from primary to secondary school. |
| Lester, 2015, Australia | Observational: prospective cohort study (4 waves between 2005 and 2007: end of primary school, beginning and end of 1st year of secondary school and end of 2nd year of secondary school) | 1800 students from 11 Catholic secondary schools in Western Australia: end of primary school (mean age 12 years, SD not available), beginning and end of 1st year of secondary school (mean age 13 years, SD not available) and end of 2nd year of secondary school (mean age 14 years, SD not available)) | Peer support: assessed with peer support scale, adapted from the Perceptions of Peer Social Support Scale | Outcomes measured: Depression, anxiety and stress, assessed with the Depression Anxiety Stress Scales-21 (DASS-21).

Statistics: Regression models |

Magro, 2019, USA	Observational: prospective cohort study (3-year study with 1-year interval)	1599 children 37 schools along the German-Dutch border, mean age at first measurement point 7.99±0.52 years	Perceived social support, assessed by the peer and family climate subscales of the German and Dutch versions of the Kid-KINDL-R	Outcome measured: Self-esteem: measured using the 4-item self-esteem subscale of the German and Dutch versions of Ravens-Sieberer's Questionnaire for Measuring Health-Related Quality of Life in Children, revised (Kid-KINDL-R). Statistics: Multilevel models
Nilsen, 2013, Norway	Observational: prospective cohort study (3 waves with 2-year intervals)	929 families (86%) participated at T1. The families were then invited to participate at six further waves (T2–T7); with the adolescents responding to questionnaires at ages 12.5 (T5), 14.5 (T6), and 16.5 (T7). Only data from T5-T7 were included in this study).	Friend support, assessed by self-report of three items measuring feelings of attachment, mutual respect, and belonging. Participants responded to items such as "I feel closely attached to my friends" with four response categories ranging from "Agree" to "Disagree".	Outcome measured: Depressive symptoms, assessed with the 13-item Short Mood and Feelings Questionnaire (SMFQ) drawn from the original 34-item Mood and Feelings Questionnaire Statistics: To test for direct effects and multiple mediation we conducted Structural Equation Modeling (SEM) with Mplus version 6. The SEM approach allows for simultaneous examination of the relative contribution of each variable, in addition to the total indirect effect, in a multiple-mediator model.
Pisarska, 2020, Poland	Observational: prospective cohort study (2 waves with a 2-year interval)	551 adolescents in the 10th and 12th grade, aged 16-18 years (mean age not available), attending public/non-public general, technical high schools and basic vocational schools from Warsaw	Friends support (Perceived Social Support from Friends (PSS-Fr)) at T1	Outcomes measured: - gambling involvement that includes both gambling-related behaviours and problems (self-report) Statistics: Generalised linear models with gamma variation
Pössel, 2018, USA	Observational: prospective cohort study (5 waves, yearly interval)	Australian students from 24 secondary schools. At baseline 2545 students (grade 8, mean age 13.1±0.5, range 11-16 years) participated and	Social support from friends, assessed using the Multidimensional Scale of Perceived Social Support (MSPSS)	Outcome measured: - depressive symptoms, assessed with the Center for Epidemiological Studies Depression Scale (CES-D).

		2405 or 94.5% of them completed the follow-up assessments in grade 9, 2219 or 87.2% in grade 10, 1717 or 67.5% in grade 11, and 1452 or 57.1% in grade 12.		Statistics: Path analyses
Rawana, 2013, Canada	Observational: prospective cohort study (2 waves with a 1-year interval)	4587 adolescents aged 12-18 years (mean age 14.95±1.55) attending a random sample of 80 high schools (sampling design of Add Health).	Perceived peer support: measured using a question that asked whether the participant felt their friends cared about them. Responses varied from 1 (very much) to 5 (not at all) and higher scores were reflective of lower perceived peer support.	Outcome measured: - depressive symptoms: assessed with a 19-item scale derived from the Centre of Epidemiologic Studies Depression Scale (CES-D) Statistics: Hierarchical regression analysis
Rosario, 2008, USA	Observational: prospective cohort study (3 waves with 1-year intervals)	Community sample of 667 middle school students, aged 11-14 years (mean age was not available), in the inner city (community violence among highest in NYC)	Social support (from peers) was assessed (at T1, T2, T3) using the 'Inventory of Parent and Peer Attachment' (Armsden and Greenberg 1987)	Outcomes measured: - anxious symptoms were assessed (at T2, T3) using the anxiety subscale of the 'Revised Children's Manifest Anxiety Scale' (RCMAS: Reynolds and Richmond 1978, 1979). - depressive symptoms during the past week were assessed (at T2, T3) using a modified version of the 'Beck Depression Inventory' (BDI; Beck et al. 1961). Statistics: Hierarchical multiple regressions of internalizing symptoms at T2 or T3
Rueger, 2008, USA	Observational: prospective cohort study (2 waves: data collection in fall after one month of school and in spring one month of the end of the school year)	636 participants from either 7th or 8th grade in a large suburban middle school (info on age of the children is lacking)	Support from classmates or close friends (Child and Adolescent Social Support Scale, CASSS) at W1	Outcomes measured at W2 with The Behavioral Assessment Scale Self Report of Personality (BASC-2 SRP) which is a 176-item rating scale that measures the personality and self-perceptions of children and adolescents. The following subscales were used as outcome measures:

				- Depression: depressed, negative affect, sadness, and loneliness -Anxiety: fear, nervousness, and worrisome behavior - Self-Esteem: global self-satisfaction Statistics: Multiple regression analyses
Slomkowski, 2001, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	The initial sample of 451 two-parent families was recruited through the cohort of all seventh-grade students (the target subjects in the study) in eight rural counties in north central Iowa who were enrolled in public or private schools during fall term, 1989. Only families that included both biological parents of the seventh graders when the project began were included in the study. 81 sister pairs and 83 brother pairs participated in all four waves of data collection. Age 9-18 year during the course of the study (mean age older sibling at baseline 13.1 years, SD not available and younger sibling 11.2, SD not available)	Warmth-support between siblings: estimated using two different sources of information: target child report of sibling's behaviors and sibling report of target's behavior	Outcomes measured: delinquency, measured with self-report questionnaire inquiring about participation in a variety of delinquent acts during the previous 12 months Statistics: Regression models to explore longitudinal patterns that are detectable in the data
Smokowski, 2016, USA	Observational: prospective cohort study (3 waves with a 1-year interval)	The final analytic sample comprised 4,065 observations at baseline, 4,251 observations at Wave 2, and 4,256 observations at Wave 3, from students 6th to 10th grade (mean age was not available) from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged	Friend support: assessed with the School Success Profile	Outcomes measured: aggression, assessed with Youth Self-Report, the adolescent version of the Child Behaviour Checklist. Statistics: An imputation model with more than 70 variables was used to fill in the missing values. Subsequent hierarchical linear model (HLM) analyses were based on the 10 and 20

		counties in North Carolina		imputed files generated in the multiple imputation.
Smokowski, 2017, USA	Observational: prospective cohort study (4 waves with a 1-year interval)	The final analytic sample for internalizing symptoms comprised 3,751 observations at baseline, 3,981 observations at Wave 2, 4,839 observations at Wave 3 and 4,216 observations at Wave 4. The final sample used in the analysis for aggressive behavior was comprised of 3,735 observations at baseline, 3,999 observations at Wave 2, 4,872 observations at Wave 3, and 4,175 observations at Wave 4. Students 6th to 11th grade from 28 public middle schools and 12 public high schools in two rural, economically disadvantaged counties in North Carolina. Mean age at baseline was 12.77±1.05 years.	Friend support: assessed with the School Success Profile	Outcomes measured: - aggression, assessed with 12-item subscale of Youth Self-Report (YSR) - internalizing symptoms, assessed with seven items from YSR that assess symptoms of anxiety and depression Statistics: Hierarchical linear model (HLM) analyses
Stice, 2004, USA	Observational: prospective cohort study (longitudinal, 3 waves with 1-year interval)	496 girls, aged 11-15 years (mean age 13±0.73 years), in public and private middle schools in a southwestern city	Perceived social support: measured with Network of Relationships Inventory (NRI)	Outcomes measured: - depressive symptoms: assessed with an adapted version of the Schedule for Affective Disorders and Schizophrenia for School-Age Children, a structured interview assessed Diagnostic and Statistical Manual of Mental Disorders major depression symptoms Statistics: Latent Growth Curve models, because this analytic technique accommodates missing data and provides a sensitive test of

				prospective relations by modeling the individual growth trajectories shown by each participant.
Swirsky, 2021, USA	Observational: Prospective cohort study (2 waves, 5-6 months apart)	321 7th grade students (mean age 13.01±0.45 years, range 12.19–14.25 years) from 3 public middle schools in an urban school district in the north-eastern USA	Peer support: reporting the amount of peer support they received	Outcomes measured: - aggression: assessed with peer nominations - loneliness: assessed with Asher Loneliness Scale Statistics: Linear regression analyses with moderation
Väänänen, 2014, Finland	Observational: prospective cohort study (2 waves with a 2-year interval)	2038 adolescents, mean age 15.5±0.39 years at T1.	Social support: assessed with the Perceived Social Support Scale-Revised (PSSS-R)	Outcomes measured: - depression: assessed with a Finnish modification of the Beck Depression Inventory - social phobia: assessed with the Social Phobia Inventory (SPIN) Statistics: Binomial logistic regression analysis
Way, 2003, USA	Observational: prospective cohort study (2 waves with a 2-year interval)	100 adolescents (mean age 14.2 at T1, SD not available) from a public high school in New York City who took part in a longitudinal study of adolescent development.	Friendship support: assesses with the Perceived Social Support Scale for Friends (PSS-FR)	Outcomes measured: - self-esteem: assessed with the Rosenberg Self-Esteem Scale (RSE) - depressive symptomatology: assessed with 10-item depressive symptom subscale of Buhrmester's measure of socioemotional adjustment. Statistics: Hierarchical multiple regression
Wentzel, 2019, USA	Observational: prospective cohort study (2 waves with a 1-year interval)	160 students, 8th grade at T1 (mean age was not available), enrolled in a mid-Atlantic region sixth-through eighth-grade middle school	Peer support: assessed with 3 items from the Peer Social Support Subscale of the Classroom Life Measure	Outcomes measured: - Emotional wellbeing, assessed with 3-item wellbeing subscale of the Weinberger Adjustment Inventory-Short Form. Statistics: Path analyses
Wright, 2019, USA	Observational: prospective cohort study (2 waves with a 1-year interval)	606 Latin adolescents from 3 middle schools, in Midwest USA. All	Social support from friends: assessed with Child and Adolescent Social Support Scale	Outcomes measured: - Self-reported aggression: adapted from the peer-nomination

		adolescents were in the 8th grade (T2), aged 13-15 years (mean age 14.36±0.46). Adolescents self-reported that their families were from Mexico (80%), Puerto Rico (10%), Guatemala (5%), and another country (5%), including Argentina, Brazil, and Paraguay		aggressive behaviours measure - peer-nominated aggression: assessed with measure by Crick & Grotpeter, 1995. 4 items assessed relational aggression, 3 items assessed physical aggression. Statistics used: Structural regression model
Young, 2005, USA	Observational: prospective cohort study (2 waves with a 2-year interval)	389 adolescents, aged 11-16 years (mean age 13.03±1.23 y) at T1 and aged 13-18 years (mean age 15.39±1.36 y) at T2 attending 100 randomly selected neighbourhoods in two counties in upstate New York	Anticipated support from peers: assessed with 6 interview questions	Outcomes measured: - depression: assessed with an unpublished modified version of the Diagnostic Interview Schedule for Children. Statistics: Linear and logistic regression analyses
Zimmerman, 2000, USA	Observational: prospective cohort study (two waves with a 6-month interval)	173 African-American male adolescents from Baltimore, Maryland, mean age at T1 16.8±1.32 years	Friend support (Shortened version of PSS-Fr) at T1	Outcomes measured: - Marijuana use with 7-point Likert scale (6 was coded as more than once a day, 0 was coded as not at all). - Depression and anxiety symptoms with subscales of the Brief Symptom Inventory Statistics: Multiple regression analysis

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
5-11 years				
Depression at T2	Close friend support	Not statistically significant: B: -0.03±0.02 £† (p>0.05)	1, 226 §	Klima, 2008
Global self-worth at T2		Not statistically significant: B: 0.08±0.07 £† (p>0.05)		
Externalising symptoms at T2		Not statistically significant: B: 0.02±0.02 £† (p>0.05)	1, 155 §	
Internalising symptoms at T2		Not statistically significant: B: -0.01±0.02 £† (p>0.05)		
Self-esteem	Peer support, within	Statistically significant: B: 0.26, 95%CI [0.21;0.31] (p<0.01) *With benefit for peer support, within*	1, 1599	Magro, 2019

	Peer support, between	Statistically significant: B: 0.31, 95%CI [0.26;0.36] (p<0.01) *With benefit for peer support, between*		
12-18 years				
Friend support				
Internalizing problem behaviour (T1 --> T3)	Friend support in older adolescents (13-15 years)	Not statistically significant: *self-report:* B: -0.01 £† (p>0.05) *parent-report:* B: -0.09 £† (p>0.05)	1, 271 §	Branje 2004
	Friend support in younger adolescents (11-13 years)	Not statistically significant: *self-report:* B: 0.06 £† (p>0.05) *parent-report:* B: 0.05 £† (p>0.05)		
Internalizing symptoms	Friend support	Not statistically significant: B: -0.008 (calculated from exp(B)) £† (p>0.05)	1, 4216	Smokowski, 2017
		Boys: Not statistically significant: B: -0.0161 £† (p>0.05)	1, 2894	Cotter, 2016
		Girls: Not statistically significant: B: -0.0010 £† (p>0.05)	1, 3000	
Externalizing problem behaviour (T1 --> T3)	Friend support in older adolescents (13-15 years)	Not statistically significant: *self-report:* B: 0.06 £† (p>0.05) *parent-report:* B: 0.03 £† (p>0.05)	1, 271 §	Branje 2004
	Friend support in younger adolescents (11-13 years)	Not statistically significant: *self-report:* B: -0.06 £† (p>0.05) *parent-report:* B: 0.05 £† (p>0.05)		
Externalising behaviour	Friend support	Boys: Not statistically significant: B: 0.0139 (calculated from exp(B)) £† (p>0.05)	1, 2894	Cotter, 2016
		Girls: Statistically significant: B: 0.0247 £ (p<0.01) *With harm for friend support*	1, 3000	
Aggression	Friend support	Not statistically significant:	1, 3580	Cotter, 2017

		B: -0.002 (calculated from exp(B)) £† (p>0.05)		
		Not statistically significant: B: 0.024 (calculated from exp(B)) £† (p>0.05)	1, 4256	Smokowski, 2016
		<u>Statistically significant</u>: B: 0.020 (calculated from exp(B)) £ (p<0.001) *With harm for friend support*	1, 1475	Smokowski, 2017
Relational aggression T2 (self-reported)	Social support from friends	<u>Statistically significant</u>: B: -0.26 £ (p<0.01) *With benefit from social support from friends*	1, 606	Wright, 2019
Relational aggression T2 (peer-nominated)		<u>Statistically significant</u>: B: -0.25 £ (p<0.01) *With benefit from social support from friends*		
Physical aggression T2 (self-reported)		Not statistically significant: B: -0.15 £† (p>0.05)		
Physical aggression T2 (peer-nominated)		Not statistically significant: B: -0.14 £† (p>0.05)		
Cyberbullying T2	Friend social support T1	Not statistically significant: B: 0.01±0.02 £† (p>0.05)	1, 1416	Fanti, 2012
Bullying T2	Friend support T1	<u>Statistically significant</u>: B: -0.10 (standardized cross-path estimate) £ (p=0.022) *With benefit for friend support*	1, 880	Kendrick, 2012
Marijuana use at T2	Friend support	Not statistically significant: β: -0.001 £† (p>0.05)	1, 173 §	Zimmerman, 2000
		Not statistically significant: β: 0.02 £† (p>0.05)	1, 166 §	Branstetter, 2011
Hard drug use at T2		Not statistically significant: β: -0.07 £† (p>0.05)		
Gambling involvement		Not statistically significant: r: -0.100 £† (univariate result, not significant, not included in multivariate model) (p>0.05) (multivariate analysis)	1, 261 §	Pisarska, 2020
Depressive symptoms	Friend support	Not statistically significant: ££† (p>0.05)	1, 1528	Pössel, 2018
		<u>Statistically significant</u>: ££† (p<0.05) *With benefit for friendship support*	1, 959	Burke, 2017
		Boys: Not statistically significant: β: 0.049 £† (p>0.05) Girls: <u>Statistically significant</u>: β: -0.234 £	1, 369-554	Nilsen, 2013

		(p<0.001) *With benefit for friend support*		
		Not statistically significant: β: -0.01 £† (p>0.05)	1, 173 §	Zimmerman, 2000
	Medium social support from friends	Not statistically significant: aOR: 0.89, 95%CI [0.42;1.87] ¥ (p>0.05)	1, 821	Khatib, 2013
	High social support from friends	Not statistically significant: aOR: 0.84, 95%CI [0.39;1.79] ¥ (p>0.05)		
Depression at T2	Friend support	Boys: Not statistically significant: B: 0.03±0.05 £† (p>0.05)	1, 313 §	Rueger, 2008
		Girls: Not statistically significant: B: -0.11±0.07 £† (p>0.05)	1, 325 §	
Anxiety symptoms at T2	Friend support	Not statistically significant: β: 0.06 £† (p>0.05)	1, 173 §	Zimmerman, 2000
Anxiety: Harm avoidance at T2		Not statistically significant: B: 0.70±3.44 £† (p>0.05)	1, 77 §	Grills-Taquechel, 2010
Anxiety: Physical symptoms at T2		Not statistically significant: B: -0.44±4.61 £† (p>0.05)		
Separation anxiety at T2		Not statistically significant: B: -4.01±3.34 £† (p>0.05)		
Social anxiety at T2		Not statistically significant: B: -3.44±3.92 £† (p>0.05)		
Anxiety at T2		Boys: Not statistically significant: B: 0.10±0.06 £† (p>0.05)	1, 313 §	Rueger, 2008
		Girls: Not statistically significant: B: 0.02±0.08 £† (p>0.05)	1, 325 §	
Psychological distress	Medium social support from friends	Not statistically significant: aOR: 0.93, 95%CI [0.58;1.49] ¥ (p>0.05)	1, 821	Khatib, 2013
	High social support from friends	Not statistically significant: aOR: 1.22, 95%CI [0.80;1.85] ¥ (p>0.05)		
Appearance esteem	Perceived Social Support from Friend	Not statistically significant: β: -0.01±0.01 £† (p>0.05)	1, 240 §	Gagné, 2020
Self-esteem	Perceived support from friends	Not statistically significant: B: 0.23±0.14 £† (p>0.05)	1, 100 §	Way, 2003
Self-esteem at T2	Friend support	Boys: Not statistically significant: B: 0.09±0.05 £† (p>0.05)	1, 313 §	Rueger, 2008
		Girls: Not statistically significant: B: -0.01±0.08 £† (p>0.05)	1, 325 §	

Peer support				
Aggression	Peer support (as mediator for overt victimization)	Not statistically significant: B: 0.033±0.0034 £† (p>0.05)	1, 321 §	Swirsky, 2021
	Peer support (as mediator for social victimization)	Not statistically significant: B: 0.043±0.0035 £† (p>0.05)		
Perpetration-victimization (beginning of grade 8)	Peer support (end of grade 7)	Statistically significant: β: -0.07 £ (p<0.05) *With benefit for peer support*	1, 1121	Lester 2012
Perpetration-victimization (end of grade 8)	Peer support (beginning of grade 8)	Statistically significant: β: -0.08 £ (p<0.05) *With benefit for peer support*	1, 1363	
Perpetration-victimization (beginning of grade 9)	Peer support (end of grade 8)	Statistically significant: β: -0.08 £ (p<0.05) *With benefit for peer support*	1, 1349	
Low risk at T1 to problematic drinking at T2 in white girls (increasers)	Peer support	Statistically significant: aOR: 1.71, 95%CI [1.45;2.02] (p<0.001) *With harm for peer support*	1, 1625	Dauber, 2011
Problematic drinking at T1 to low risk at T2 in white girls (decliners)		Statistically significant: aOR: 1.90, 95%CI [1.48;2.43] (p<0.001) *With benefit for peer support*		
Low risk at T1 to problematic drinking at T2 in African American girls (increasers)		Not statistically significant: aOR: 1.11, 95%CI [0.78;1.60] ¥ (p>0.05)	1, 600	
Problematic drinking at T1 to low risk at T2 in African American girls (decliners)		Not statistically significant: aOR: 1.16, 95%CI [0.72;1.88] ¥ (p>0.05)		
Depressive symptoms	Peer support	Not statistically significant: β: -1.88±1.62 £† (p>0.05)	1, 186 §	Arora, 2017
	Perceived support from friends	Not statistically significant: B: -0.03±0.03 £† (p>0.05)	1, 100 §	Way, 2003
	Anticipated peer support at T1	Not statistically significant: B: 0.55±1.11 £† (p>0.05)	1, 389 §	Young, 2005
Depressive symptoms (in participants trying to lose weight)	Perceived peer support	Statistically significant: β: 0.03±0.11 £ (p≤0.01) *With benefit for peer support*	1, 4587	Rawana, 2013
Depressive symptoms (in participants engaged in weight loss behaviors)		Statistically significant: β: 0.04±0.11 £ (p≤0.01) *With benefit for peer support*		
Depressive symptoms (in participants with weight perceptions)		Statistically significant: β: 0.03±0.11 £ (p≤0.01) *With benefit for peer support*		
Increase in depressive symptoms (T1>T3)	Peer social support	Not statistically significant: B: -0.009, 95%CI [-0.033;0.015] £†	1, 496	Stice, 2004

		(p=0.445)		
Depressive symptoms (T1 to T2 change --> T2)	Social support (from peers)	Boys: <u>Statistically significant:</u> β: -0.26 £ (p<0.001) *With benefit for social support (from peers)* Girls: Not statistically significant: β: -0.09 £† (p>0.05)	1, 677	Rosario, 2008
Depressive symptoms (T1 to T3 change --> T3)	Social support (from peers)	Boys: <u>Statistically significant:</u> β: -0.22 £ (p<0.01) *With benefit for social support (from peers)* Girls: Not statistically significant: β: -0.09 £† (p>0.05)		
Depressive symptoms in 2-year follow-up period	Peer support	<u>Statistically significant:</u> B: -0.03, 95%CI [-0.05;-0.01] (p=0.009) *With benefit for peer support*	1, 496	Burton, 2004
Depressive symptoms at T2		<u>Statistically significant:</u> β: -0.13 £ (p<0.05) *With benefit for peer support*	1, 217 §	Colarossi, 2003
Depressive symptoms at T5		Not statistically significant: B: -0.28±0.46 £† (p>0.05)	1, 258 §	Auerbach, 2011
Major depression at T2	Anticipated peer support at T1	Not statistically significant: aOR: 0.56, 95%CI [0.18;1.73] ¥ (p>0.05)	1, 389 §	Young, 2005
Depression at T2	Girl's perceived social support at T1	<u>Statistically significant:</u> aOR: 0.897, 95%CI [0.812;0.990]$ (p=0.031) *With benefit for perceived social support*	1, 30 vs 821	Väänänen, 2014
Depression at T2	Boy's perceived social support at T1	Not statistically significant: aOR: 0.933, 95%CI [0.832;1.047]$ (p=0.238)	1, 14 vs 672	Väänänen, 2014
Depression	Peer support	*End of 1st year of secondary school:* Not statistically significant: β: -0.50, 95%CI [-1.80;0.80] £† (p>0.05) *End of 2nd year of secondary school:* <u>Statistically significant:</u> β: -2.85, 95%CI [-4.02;-1.68] (p<0.01) *With benefit for peer support*	1, 1800	Lester, 2015
Onset of subthreshold and threshold depression		Not statistically significant: B: -0.30, 95%CI [-0.70;0.10] £† (p=0.143)	1, 496	Burton, 2004
Depression severity		Not statistically significant: B: -0.14, 95%CI [-0.29;0.02] £†	1, 368 §	Jaycox, 2009

		(p=0.086)		
Anxiety		*End of 1st year of secondary school:* Not statistically significant: β: -0.66, 95%CI [-1.73;0.43] £† (p>0.05) *End of 2nd year of secondary school:* Statistically significant: β: -2.90, 95%CI [-3.89;-1.92] (p<0.01) *With benefit for peer support*	1, 1800	Lester, 2015
Anxiety symptoms (T1 to T2 change --> T2)	Social support (from peers)	Boys: Statistically significant: β: -0.20 £ (p<0.01) *With benefit for social support (from peers)* Girls: Not statistically significant: β: -0.07 £† (p>0.05)	1, 677	Rosario, 2008
Anxiety symptoms (T1 to T3 change --> T3)		Boys: Statistically significant: β: -0.16 £ (p<0.05) *With benefit for social support (from peers)* Girls: Statistically significant: β: -0.17 £ (p<0.01) *With benefit for social support (from peers)*		
Social anxiety at T3	Peer support	Statistically significant: B: -0.28 £ (p<0.05) *With benefit for peer support*	1, 416	Cavanaugh, 2015
Anxious symptoms at T5		Not statistically significant: B: 0.13±0.15 £† (p>0.05)	1, 258 §	Auerbach, 2011
Stress		*End of 1st year of secondary school:* Not statistically significant: β: -0.51, 95%CI [-1.78;0.77] £† (p>0.05) *End of 2nd year of secondary school:* Statistically significant: β: -0.39, 95%CI [0.32;0.42] (p<0.01) *With harm for peer support*	1, 1800	Lester, 2015
Dependent interpersonal stress at T5		Not statistically significant: B: -1.37±0.87 £† (p>0.05)	1, 258 §	Auerbach, 2011
Self-esteem at T2		Statistically significant: β: 0.14 £ (p<0.05) *With benefit for peer support*	1, 217 §	Colarossi, 2003

Loneliness	Peer support (as mediator for overt victimization)	Statistically significant: B: -0.156±0.0031 (p<0.01) *With benefit for peer support*	1, 321 §	Swirsky, 2021
	Peer support (as mediator for social victimization)	Statistically significant: B: -0.154±0.0031 (p<0.01) *With benefit for peer support*		
	Peer support	Not statistically significant: B: 0.06 £† (p>0.05)	1, 416	Cavanaugh, 2015
Wellbeing (grade 9)	Perceived support from peers (grade 8, indirect effect)	Statistically significant: B: 0.07, 95%CI [0.01;0.15] (p<0.05) *With benefit for perceived support from peers*	1, 160 §	Wentzel, 2019
	Perceived support from peers (grade 8, direct effect)	Not statistically significant: B: -0.07 £† (p=0.48) *With benefit for perceived support from peers*		
Social phobia at T2	Girl's perceived social support at T1	Not statistically significant: aOR: 0.933, 95%CI [0.855;1.017] $ (p=0.114)	1, 41 vs 810	Väänänen, 2014
	Boy's perceived social support at T1	Not statistically significant: aOR: 0.945, 95%CI [0.870;1.026] $ (p=0.176)	1, 30 vs 656	
School satisfaction	Peer support for learning	Not statistically significant: B: 0.060 £† (p>0.05)	1, 614 §	Jiang, 2012
Eating disorder vs Low ChEAT (mean±SE)	Conditional Support by Peer	Not statistically significant: 16.08±0.74 vs 14.75±0.80 $ MD: 1.33 £† (p>0.05)	1, 25 vs 25 §	Kirsch, 2007
Eating disorder vs High ChEAT (mean±SE)		Not statistically significant: 16.08±0.74 vs 15.32±0.81 $ MD: 0.76 £† (p>0.05)		
Sibling support				
Externalizing problem behaviour (T1 --> T3)	Sibling support in older adolescents (13-15 years)	Not statistically significant: *self-report:* B: -0.05 £† (p>0.05) *parent-report:* B: 0.01 £† (p>0.05)	1, 271 §	Branje 2004
	Sibling support in younger adolescents (11-13 years)	Not statistically significant: *self-report:* B: -0.15 £† (p>0.05) *parent-report:* B: 0.09 £† (p>0.05)		
Internalizing problem behaviour (T1 --> T3)	Sibling support in older adolescents (13-15 years)	Not statistically significant: *self-report:* B: -0.10 £† (p>0.05) *parent-report:* B: 0.12 £† (p>0.05)		

	Sibling support in younger adolescents (11-13 years)	Not statistically significant: *self-report:* B: -0.10 £† ($p>0.05$) *parent-report:* B: 0.11 £† ($p>0.05$)		
Younger sister's delinquency	Warmth-support between siblings	Statistically significant: β: -0.37 £ ($p<0.001$) *With benefit for warmth-support between siblings*	1, 81 §	Slomkoswki, 2001
Younger brother's delinquency		Not statistically significant: β: 0.31 £† ($p<0.01$) *With harm for warmth-support between siblings*	1, 83 §	
Classmate support				
Depression at T2	Classmate support	Boys: Statistically significant: B: -0.28±0.06 ($p<0.01$) *With benefit for classmate support* Girls: Not statistically significant: B: -0.03±0.06 £† ($p>0.05$)	1, 313 § 1, 325 §	Rueger, 2008
Depressive symptoms at T5		Statistically significant: B: 1.73±0.46 ($p<0.001$) *With harm for poor classmate support*	1, 258 §	Auerbach, 2011
Anxiety: Harm avoidance at T2		Not statistically significant: B: 1.76±1.55 £† ($p>0.05$)	1, 77 §	Grills-Taquechel, 2010
Anxiety: Physical symptoms at T2		Not statistically significant: B: 1.21±2.05 £† ($p>0.05$)		
Separation anxiety at T2		Not statistically significant: B: 0.90±1.51 £† ($p>0.05$)		
Social anxiety at T2		Not statistically significant: B: 2.67±1.82 £† ($p>0.05$)		
Anxiety at T2		Boys: Statistically significant: B: -0.30±0.07 ($p<0.01$) *With benefit for classmate support* Girls: Not statistically significant: B: 0.01±0.08 £† ($p>0.05$)	1, 313 § 1, 325 §	Rueger, 2008
Anxious symptoms at T5		Not statistically significant: B: -0.01±0.16 £† ($p<0.001$)	1, 258 §	Auerbach, 2011

Self-esteem at T2		Boys: _Statistically significant_: B: 0.13±0.06 (p<0.05) _With benefit for classmate support_	1, 313 §	Rueger, 2008
		Girls: Not statistically significant: B: 0.05±0.08 £† (p>0.05)	1, 325 §	
Dependent interpersonal stress at T5		_Statistically significant_: B: -4.42±0.82 (p<0.001) _With benefit for classmate support_	1, 258 §	Auerbach, 2011

Mean ± SD (unless otherwise indicated), B ± SE (unless otherwise indicated), MD: mean difference, aOR: adjusted odds ratio, SD: standard deviation, SE: standard error, B: unstandardized coefficient, β: standardized coefficient

$ The outcome measures and effect measures represent the risk factor/exposure, not the outcome

£ No SE and/or CI available; or no information on magnitude of effect available to assess variability of results

££ No raw data or effect size reported

¥ Imprecision (large variability of results)

† Imprecision (lack of data)

§ Imprecision (limited sample size or low number of events)

Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Arora, 2017	Unclear, little information available on recruitment process (part of larger study). Small subset of the total sample, only 1%, participated in this study.	Yes, exposure and outcome measures were self-reported.	No, T1 moderator variables (i.e., parent support, peer support, teacher support, school engagement) and covariates (gender, grade, …) were added to the model with T2 depressive symptoms as the dependent variable.	No, attrition analyses between youth who completed the assessment at T2 and those who did not suggested no differences on demographic and study variables.	N/A
Auerbach, 2011	Unclear, little if any information on how schools and adolescents were recruited. No overall response rate was reported.	Yes, self-report measures in order to assess social support, depressive symptoms, and anxious symptoms	No, baseline depressive and anxious symptoms, age, and gender were entered in the first step of all analyses as covariates.	Unclear, the average rate of retention for each follow-up during the course of the study was 82%, and each participant completed at least three assessments. While the present sample is in line with the greater demographic distribution of Quebec, it is not possible to	N/A

				determine if there were other differences between the participants and those who opted out of the present study.	
Branje 2004	No, families were selected from civil registry lists provided by 23 randomly selected municipalities across the Netherlands	No, use of questionnaires, but both self-reported as parent-reported outcomes	No, step-wise hierarchical regression analysis performed, adjusted for age, sex, number of siblings.	No, 285 of the initial 288 families still participated in wave 3.	N/A
Branstetter, 2011	Unclear, no information on recruitment procedure (probability or non-probability sampling?)	Yes, outcome measures were self-reported. Adolescents may underreport their substance usage. However, multiple reporters (friends and mothers) to assess qualities of relationships were used.	No, hierarchical regression analyses were conducted in order to determine how friendship factors and mother-adolescent relationship quality operate together in predicting the frequency of substance use. To control for prior use, substance use frequency in the 10th grade was entered as the first step (baseline result).	No, no indications of loss to follow-up	Yes, the present study focused on a single close friendship. It would also be important to incorporate other key relationships that emerge in adolescence, including other-sex friends and romantic partners.
Burke, 2017	No, students were selected through a process whereby four schools from each of the cantons with at least three classrooms were randomly selected. 75.5% of all students participated in all four waves of assessment.	Yes, the reliance on solely student self-reports may have resulted in an underreporting of depressive symptoms, to avoid stigmatization associated with being a victim. The use of only a limited number of the original items on the support, and depressive scales may have also adversely influenced the findings of the study.	No, trivariate cross-legged models were used to investigate the moderating effect of emotional support on the association between victimization and depressive symptoms.	No, attrition between the points of assessment was thus very low and was mainly due to participants relocating to another school. Accordingly, it was assumed that data were missing at random, and the full information maximum likelihood (FIML) method was used to address missing data.	N/A

Burton, 2004	Unclear, info on selection process of adolescent girls is largely lacking. The recruitment rate 56% was relatively low and self-selection or selection by parents might influence the results.	Unclear, exposures relied on self-report measures but outcome was assessed with structured diagnostic interviews.	No, regression model to adjust for other exposures (limited)	No, attrition analyses verified that girls who dropped from the study did not differ significantly from the remaining girls in terms of age, ethnicity, parental education, parental social support, peer social support, negative life events, or depressive symptoms at T1.	N/A
Cavanaugh, 2015	Unclear, self-selection possible through letter inviting but preliminary analyses indicated that eligible participating families were similar to eligible non-participating families on all study variables, suggesting minimal selection bias. The response rate was 56.8% for children and 36.8% for the families.	Yes, exposure and outcome measures were self-reported. Questionnaires also were mailed home and family members were asked to complete their booklets independently but not used for the factors and outcomes of interest.	No, structural equation modelling to correct for factors measured	No, 82% retention rate of W1 families but attrition analyses using multivariate analysis of variance were conducted using the W1 data, and there were no differences between the retained and attrited families on any of the study variables.	N/A
Colarossi, 2003	Unclear, study is based on other, larger longitudinal study but process of recruitment and selection is largely unknown. Response rate of about 60% (217 students completed the survey at both times).	Yes, exposure and outcome measures were self-reported.	No, structural equation modelling to correct for factors measured	No, attrition analyses were done and there were no significant differences in demographic variables or in levels of support, depression or self-esteem between students who were used in the analyses and those who were not.	N/A
Cotter, 2016	No, in County 1, all middle school students in 6th,	Yes, the study relied	No, model was developed and	Unclear, missing data ranged from	N/A

			0.19 to 0.35 from W1 to W4 and was not completely at random and all the analysis variables were used in multiple imputation to ensure the representative ness of covariance structure in the imputed data.	
Cotter 2017	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40 % of middle-school students was included.	Yes, the study relied exclusively on self-report measures and therefore represents adolescents' individual perceptions.	No, all variables except for demographics and indirect aggression (which was only measured at Wave 4) were included as time varying covariates.	Unclear, in each wave, additional students entered the study.	Given the unique racial/ethnic makeup of the rural sample, results should be cautiously generalized to other settings.
Drauber, 2011	No, random sample and weighting factors used to correct for unequal probabilities. 66% of the sample completed both study waves.	Yes, exposure and outcome measures were self-reported. Adolescents may underreport their substance usage.	No, the following potential W1 predictors were entered simultaneously into each regression: school connectedness, academic misbehavior, peer support, family relationship quality, depressive symptoms, delinquency, and other substance use. All models controlled for adolescent age, mother education, and family composition (single parent vs. two-parent).	Yes, 723 girls lost to follow-up which were significantly older than the sample group. No other differences	N/A
Fanti 2012	No, middle schools randomly selected	Yes, exposure and outcome were self-reported.	No, controlled for gender, age, cyberbullying, cybervictimization, school-bullying and school victimization	No, 6.41% attrition due to inability to contact students who had moved away or transferred to a different school	N/A

Gagné 2020	No, participants came from 2 high schools of similar socio-economic backgrounds and were randomly assigned to either to the experimental condition or the control group. To achieve the aim of the current study, only the students from the control group was selected.	Yes, exposure and outcome were self-reported.	No, gender, students' grade level, current perceived figure, and negative weight-related comments were included as control variables.	No, the logistic regression showed that dropout variation was not related to any of the variables included in the current study.	N/A
Grills-Taquechel, 2010	Yes, children only recruited in a single school and no further info on selection.\n\nResponse rate of only 27.5%	Yes, exposure and outcome measures were self-reported.	No, hierarchical cluster models were used to adjust for other variables of interest.	No, no significant differences on any demographic variable of interest were found between students who completed only T1 data and those who completed both assessment points.	N/A
Jaycox, 2009	Yes, diverse set of settings but no random selection and not all teens who were approached enrolled in the study, and selection bias might have affected the results.	Yes, all measures were based on self-reports.	No, to control for demographic characteristics, the authors residualized the depression, functioning, and coexisting problems variables with respect to age, gender, ethnicity, income, and insurance status; and these residualized variables were used in subsequent crosslagged models.	Unclear, 89% completed the follow-up questionnaire but no attrition analysis performed.	N/A
Jiang 2012	Yes, only students from one school were selected.	Yes, exposure and outcome were self-reported.	No, 2 sets of hierarchical regression models were tested to examine the relationships between the three school-related social support variables and school satisfaction at two different points in time.	Unclear, only data from students who completed the full surveys at both timepoints were included. Unclear if there were significant differences between these students and the ones who did not (fully)	N/A

				complete the survey at both timepoints.	
Khatib, 2013	No, 30 out of 42 secondary schools in East London were randomly selected and balanced to represent single and mixed sex schools. Eligible schools were stratified by borough and school type.

At baseline, the overall response rate was 84% and, at follow-up, 78%. | Yes, all measures were based on self-reports. It was not possible to obtain parent or teacher ratings of the SDQ which is a weakness of the study. | No, logistic regression models were used to adjust for age, gender, an interaction between age and gender, socio-economic status, SDQ as a continuous score at Phase W1, country of birth and length of time in the UK. | Unclear, no attrition analysis performed but logistic regression models, with appropriate weights for the clustered design and non-response at follow-up were used. The potential bias introduced by missing outcome data was assessed by post-hoc sensitivity analyses. | N/A |
Kendrick 2012	Unclear, the sample included participants from an ongoing longitudinal study within seven junior high schools (grades 7 through 9) in a town in central Sweden with a population of around 130,000.	Yes, exposure and outcome were self-reported.	No, a design was used that controlled both for stability over time and for cross-sectional intercorrelations of all variables.	No, 108 of 988 participants did not have data for both timepoints. Only students who participated at both timepoints were included. The group of 108 that was removed did not differ significantly from the participants in the analytical sample.	N/A
Kirsch 2007	No, cases and controls were age matched.	Yes, exposure was self-reported.	No, controlling for body mass index (BMI), multivariate analyses of variance (MANOVA) using group (ED group, community sample with low ChEAT scores, community sample with high ChEAT scores) as the between-group variable were conducted on the following groups of variables.	No, no missing data	N/A
Klima, 2008	Yes, 3 different school were selected based on convenience sampling and no info on how	Yes, close friend support, depressive symptoms and self-worth were only based on	No, hierarchical multiple regression analyses tested longitudinal associations	No, good retention rate for self-reports and, when comparing children who	N/A

	children were recruited. At baseline, the overall response rate was 36.5% and, at follow up, 91.5% for self-reports and 63.2% for teacher reports.	self-reports. However, internalising and externalising symptoms were reported by teachers.	between the two indicators of peer relations and the measures of psychological adjustment. Each regression model included as a control the T1 score that children received on the corresponding criterion. Moderation of the above associations by gender was tested as well.	did and did not have teacher data in 6th grade, no differences were found on any of the 4th grade variables.	
Lester 2012	No, schools were stratified according to the total number of students enrolled at the school and each school's socio-economic status (SES) and were randomly selected.	Yes, exposure and outcome were self-reported.	No, previous bullying involvement, gender, study condition (to control for any possible intervention effects) and clustering at the school level were taken into account in all models.	Yes, parental consent was provided for 3462 of the 3769 (92%) students eligible to participate with 3123 (90%) of the students involved in the SSP study responding to at least 3 of the 4 data collection points and 1771 responding to all 4 data points (51%).	N/A
Lester 2015	No, schools were stratified according to the total number of students enrolled at the school and each school's SES and were randomly selected.	Yes, exposure and outcome were self-reported.	No, gender and previous school climate measures were taken into account.	Unclear, only the comparison groups of an experimental study were enrolled, but unclear if there were any missing data between enrolment and the last dataset.	N/A
Magro 2019	Unclear, not clear if schools were randomly selected.	Yes, exposure and outcome were self-reported.	No, adjusted for family migration background and parent education level.	No, no significant differences between students who participated in all timepoints, compared to those who only participated in one or 2 timepoints.	N/A
Nilsen 2013	Unclear, families from 19	Yes, exposure and outcome	No, structural equation modeling	No, 34% dropouts	N/A

	geographic health care areas were invited to complete a survey when visiting a public health clinic for their scheduled 18-month vaccination for the index child. Routinely, more than 90% of all Norwegian families with children attend a public health program during the first four years of the child's life.	were self-reported.	used, which allows for simultaneous examination of the relative contribution of each variable, in addition to the total indirect effect, in a multiple-mediator model.	between t5 and t7 (data used in the study), but attrition analyses from baseline to t7 showed that only low maternal educational level, not other variables (such as mother's temperament and psychological distress, child's temperament, and mothers' emotional support from partner and friends) predicted drop-out.	
Pisarska, 2020	No, random sampling of clusters (class rooms) Response rate was about 83% for W1.	Yes, exposure and outcome measures were self-reported. These data are burdened with recall and social desirability bias.	No, model to control for demographics and W1 gambling involvement	Yes, only 65% response rate from W1 to W2. Youth omitted because of missing data may have on average a higher Gambling Involvement Index score. This selection of students in longitudinal analyses may have biased the sample towards lower risk students in gambling involvement.	N/A
Pössel 2018	Unclear, control group of experimental study used for data analysis.	Yes, exposure and outcome were self-reported.	No, significant factors taken into account in models for path analysis	Yes, non-random attrition rate of 46.63% between grade 8 and grade 12. Those who left their baseline school reported less family, friend, and teacher support at baseline, more stressful events, a higher	N/A

				frequency of depressive symptoms, and/or were of Torres Strait Islander or Aboriginal background.	
Rawana 2013	No, data from large Add Health consortium. Random sample of 80 high schools.	Yes, exposure and outcome were self-reported.	No, stepwise hierarchical regression analysis used, with gender and depressive symptoms at baseline as covariates	Yes, response rate in Wave 1 = 79%, in Wave 2 = 88.6%. No information on differences between dropouts and participants	N/A
Rosario, 2008	Yes, no random sampling	Yes, exposure and outcome were self-reported.	No, multivariable analysis accounted for baseline symptoms of depression and anxiety and for potential confounders: e.g. social desirability, victimization, witnessing of family violence.	No, retention rates were 92% at T2 and 87% at T3, with 85% of the sample interviewed at all 3 rounds. Importantly, only 7% was lost to both follow-up rounds.	N/A
Rueger, 2008	Yes, recruitment in a single school (convenience sample) The sample comprised about 74% of the total school body.	Yes, self-report assessments of both predictor and outcome variables.	No, regression analyses were used in which was adjusted for different sources of support.	Unclear, no attrition analyses were performed and it is unknown if there were any differences between the included and missing groups.	N/A
Slomkowski 2001	No, cohort of all seventh-grade students, male and female, in 8 rural counties in north central Iowa who were enrolled in public or private schools during fall term, 1989. Only families that included both biological parents of the seventh graders when the project began were included in the study.	Yes, exposure and loneliness outcome were self-reported. Aggression was measured by peer nomination.	No, step-wise hierarchical regression analysis performed	No, no significant differences between continuing and drop-out families in family income, emotional distress of parents, or levels of delinquent behavior.	N/A
Smokowski 2016	No, in County 1, all middle school students in 6th, 7th, and 8th	Yes, exposure and outcome were self-reported.	No, model was developed and adjusted for covariates	No, multiple imputation analyses were completed to	Given the unique racial/ethnic makeup of

			(demographics, positive proximal processes, and negative proximal processes (measured at baseline))	minimize the impact of missing data, which could occur due to participant non-response to questions or due to attrition from the sample over time. An imputation model with more than 70 variables was used to fill in the missing values. Subsequent HLM analyses were based on the 10 and 20 imputed files generated in the multiple imputation.	the rural sample, results should be cautiously generalized to other settings.
	grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40 % of middle-school students was included.				
Smokowski 2017	No, in County 1, all middle school students in 6th, 7th, and 8th grades were included in the sample. Because County 2 was geographically bigger with a larger student population, a random sample of 40 % of middle-school students was included.	Yes, exposure and outcome were self-reported.	No, model was developed and adjusted for covariates (15 predictors which can be categorized into the following three categories: (a) demographics, (b) microsystem protective factors, and (c) microsystem risk factors.	No, multiple imputation analyses were completed to minimize the impact of missing data.	Given the unique racial/ethnic makeup of the rural sample, results should be cautiously generalized to other settings.
Stice 2004	Unclear, no information on how girls were recruited	Yes, exposure and outcomes were self-reported (survey + interview).	No, as preliminary analyses indicated that age, ethnicity, and parental education were not significantly related to changes in parental support, peer support, or depressive symptoms over time, these demographic factors were not included as covariates in the models. There was no evidence of quadratic effects for the independent variables.	No, of the initial 496 girls, 10 did not provide T2 data and 10 did not provide T3 data, but only 4 did not provide both T2 and T3 data. Attrition analyses verified that girls who dropped from the study did not differ from the remaining girls on age, ethnicity, parental education,	N/A

				parental support, peer support, or depressive symptoms at T1.	
Swirsky 2021	Unclear, students included from 3 schools, but no information if the students/ schools were randomly selected.	Yes, exposure and outcome were self-reported.	No, linear regression analyses with moderation were conducted separately for each form of victimization (social and overt). Each model contained both forms of victimization (social and overt), the control variables (sex, race, lunch status, and the Time 1 level of the outcome variable), the Time 2 level of the peer moderators (peer support, peer preference, social status).	Unclear, no information on missing data	N/A
Väänänen 2014	Unclear, study conducted in 2 Finnish cities that represent well the Finnish urban population. No information on random selection of participants.	Yes, exposure and outcome were self-reported.	No, controlled covariates were age, family structure, parents' highest educational qualifications and externalizing symptoms.	Yes, there were more subjects with depression among dropouts. There were no significant differences between responders and dropouts for support from family among girls.	Lack of control for other possible disorders
Way 2003	Yes, adolescents from only one school included.	Yes, exposure and outcome were self-reported.	No, hierarchical multiple regression analysis was conducted to explore the independent and combined effects of demographic variables (ethnicity and gender) and contextual variables (perceptions of family support, friendship support, and school climate).	No, there were no significant demographic differences or differences in reported levels of family support, friend support, perceived school climate, or psychological adjustment at Time 1 between those who were retained for the study and those who were not.	N/A
Wentzel 2019	Unclear, no information on	Yes, exposure and outcome	No, path model used. The model featured 4	No, 160 students who enrolled at T1	N/A

	how students were recruited.	were self-reported.	simultaneous mediation models: the indirect effects of perceived emotional support from teachers and peers, unknown control beliefs, and perceived competence at Grade 8 on emotional well-being at Grade 9 were assessed. Well-being at Grade 8 was also included as a control variable.	also participated at T2.	
Wright 2019	Yes, schools were selected based on number of Latinx adolescents and their desire to participate.	Yes, exposure was self-reported. Outcome was both self-reported as by peer-nomination.	No, controlled for Time 1 peer-nominated and self-reported relational aggression and physical aggression	No, only students who participated at both timepoints were included in the analysis.	
Young 2005	No, adolescents were cluster sampled from 100 randomly selected neighbourhoods in 2 counties in upstate New York.	Yes, exposure and outcome were self-reported.	No, linear and logistic regression analysis performed. Main effects (gender, age, parent support, anticipated peer support) and the interaction of parent and peer support were included in the final model.	Yes, 29 adolescents did not complete T2. They had significantly lower mean depression scores than the remaining 389 adolescents and were significantly more likely to come from families of lower socio-economic status.	The measure of anticipated peer support, which was a composite dichotomous variable, might have impacted the findings. This measure, which emphasizes judgments about one's peer network rather than actual support from peers, is not as strong as peer support measures used in other studies.
Zimmerman 2000	Unclear, broad recruitment but probability sampling only in school-district dropout lists. Risk of self-selection in other sampling strategies. Little if any recruitment effect was detected in the analysis.	Yes, exposure and outcome measures were collected through interview procedures. These data are burdened with recall, social desirability and interviewer bias. Authors had foreseen strategies to	No, multivariate model was used to control for parent support, stressful events, interaction effects and longitudinal effects (baseline results).	Yes, only 68% response rate from W1 to W2 and youths who completed both interviews reported fewer stressful life events and less delinquency than youths who completed only the first interview.	N/A

		minimize this bias as good as possible.			

Certainty of the body of evidence

<u>5-11 years</u>

Friend support	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Peer support	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	0	
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

<u>12-18 years</u>

Friend support	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Peer support	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Limited sample sizes/lack of data/large variability of results
Inconsistency	-1	For some outcomes conflicting results were found
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	Final grading Very low [D]	

Sibling support	Initial grading Low [C]	Downgrading due to

Limitations of study design	0	See table 'Study limitations'
Imprecision	-1	Low sample size/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Classmate support	**Initial grading Low [C]**	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Low sample size/lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	**5-11 years** *Friend support* There is limited evidence concerning the risk of psychological outcomes in presence of close ***friend support***. A statistically significant change in depression, global self-worth, externalizing symptoms or internalizing symptoms in presence of close friend support could not be demonstrated (Klima 2008). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. *Peer support* There is limited evidence with benefit for ***peer support*** at both the within-person (i.e., individuals who report changes in social support also report changes in their self-esteem) and between-person (i.e., individuals who tend to have stronger relationships also tend to have stronger self-esteem levels) level. Separating and comparing within- and between-person effects is crucial to differentiate among intraindividual and interindividual effects of support. It was shown that peer support (within and between) resulted in a statistically significant increase in self-esteem (Magro 2019). Evidence is of very low certainty. **12-18 years** *Friend support* There is limited evidence <u>with benefit</u> from friend support. It was shown that friend support resulted in a statistically significant decrease in **bullying** (Kendrick 2012) and **depressive symptoms** (Burke 2017, Nilsen 2013 (girls only)). However, a statistically significant decrease in cyberbullying in presence of friend support could not be demonstrated (Fanti, 2012). Moreover, statistically significant decrease in depressive symptoms in presence of friend support could not be demonstrated in four studies (Nilsen 2013, Khatib 2013, Pössel 2018, Rueger 2008, Zimmerman 2000). Evidence is of very low certainty and results cannot be considered precise due to lack of data. There is <u>limited evidence</u> concerning the risk or presence of **general mental health problems** (internalizing or externalizing problem behaviour, psychological distress or self-esteem), **addiction or anxiety,** in presence of *friend support*.

A statistically significant change in internalizing problem behaviour, psychological distress, appearance esteem or self-esteem in presence of friend support could not be demonstrated (Branje 2004, Cotter 2016, Gagné 2020, Khatib 2013, Smokowski 2017, Way 2003). Also, a statistically significant decrease in externalizing problem behaviour in presence of friend support in younger or older adolescents (Branje 2004) or friend support in boys (Cotter 2016) could not be demonstrated. However, it was shown in one study that friend support in girls resulted in a statistically significant increase in externalising behaviour (Cotter 2016). *Please see the Dutch summary for possible explanations on this latter counterintuitive result.*
A statistically significant decrease in marijuana use, hard drug use or gambling involvement, or anxiety (harm avoidance, physical symptoms of anxiety, separation anxiety or social anxiety), in presence of friend support could not be demonstrated (Branstetter 2011, Grills-Taquechel 2010, Pisarska 2020, Rueger 2008, Zimmerman 2000).

Evidence is of very low certainty and results cannot be considered precise due to limited sample size, large variability of results and/or lack of data.

There is conflicting evidence concerning the risk of **aggression** in presence of friend support.
On the one hand, it was shown that friend support resulted in a statistically significant increase in aggression (Smokowski 2017). On the other hand, it was shown that friend support resulted in a statistically significant decrease in relational aggression (both self-reported as peer-nominated) (Wright 2019). *Please see the Dutch summary for possible explanations on these conflicting results.*
Moreover, a statistically significant decrease in (physical) aggression in presence of friend support could not be demonstrated in 3 studies (Cotter 2017, Smokowski 2016, Wright 2019).
Evidence is of very low certainty and results cannot be considered precise due to lack of data.

Peer support
There is limited evidence with benefit for peer support.
It was shown that peer support resulted in a statistically significant decrease in **perpetration-victimization** (Lester 2012). However, a statistically significant decrease in aggression in presence of peer support could not be demonstrated (Swirsky 2021).
It was shown in 6 studies that peer support resulted in a statistically significant decrease in **depressive symptoms** (Burton 2004, Colarossi 2003, Lester 2015, Rawana 2013, Rosario 2008 (boys only), Väänänen 2014 (girls only))
However, a statistically significant decrease in depressive symptoms or onset of depression in presence of peer support could not be demonstrated in 9 tudies (Arora 2017, Auerbach 2011, Burton 2004, Jaycox 2009, Rosario 2008 (girls only), Stice 2004, Väänänen 2014 (boys only), Young 2005, Way 2003).
It was shown in 3 studies that peer support resulted in a statistically significant decrease in **anxiety** (Cavanaugh 2015, Lester 2015 (end of 2nd year), Rosario 2008 (boys only)). However, a statistically significant decrease in anxiety in presence of peer support could not be demonstrated in 2 studies (Auerbach 2011, Lester 2015 (end of 1st year)).
It was shown that peer support resulted in a statistically significant increase in **self-esteem** (Colarossi 2003).
It was shown that peer result resulted in a statistically significant decrease in **loneliness** (Swirsky 2021). However, a statistically significant decrease in loneliness in presence of peer support could not be demonstrated in another study (Cavanaugh 2015).
It was shown that perceived support from peers resulted in a statistically significant increase in **wellbeing** (Wentzel 2019).
Evidence is of very low certainty and results cannot be considered precise due to limited sample size and/or lack of data.

There is limited evidence with harm for peer support for the outcome stress.
It was shown that peer support resulted in a statistically significant increase in **stress** at the end of 2nd year of secondary school. However, a statistically significant change in stress at the end of 1st year of secondary school could not be demonstrated (Lester 2015). In addition, a statistically significant change in

	dependent interpersonal stress in presence of peer support could not be demonstrated (Auerbach 2011). Evidence is of very low certainty and results cannot be considered precise due to limited sample size, large variability of results and/or lack of data. There is limited evidence concerning the risk of **social phobia, school satisfaction or eating disorders** in presence of *peer support.* A statistically significant decrease in social phobia or eating disorders in presence of peer support, could not be demonstrated (Jiang 2012, Kirsch 2007, Väänänen 2014). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. There is <u>conflicting evidence</u> from one observational study concerning **problematic drinking in white girls** in presence of peer support. On the one hand, it was shown that peer support resulted in a statistically significant decrease in problematic drinking increasers (Dauber 2011). On the other hand, it was shown that peer support resulted in a statistically significant increase in problematic drinking decliners (Dauber 2011). *Please see the Dutch summary for possible explanations on these conflicting results.* However, a statistically significant change in problematic drinking in African American girls in presence of peer support could not be demonstrated (Dauber 2011). Evidence is of very low certainty and results cannot be considered precise due to limited sample size, large variability of results and/or lack of data. *Sibling support* There is limited evidence concerning the risk of **internalizing or externalizing problem behaviour** in presence of sibling support. A statistically significant decrease in externalizing or internalizing problem behaviour in presence of sibling support in younger or older adolescents could not be demonstrated (Branje 2004). However, there is limited evidence with benefit from warmth *support between siblings*. It was shown that warmth-support between siblings resulted in a statistically significant decrease in delinquency (Slomkowski 2001). Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data. *Classmate support* There is limited evidence with benefit from classmate support. It was shown that classmate support resulted in a statistically significant decrease in **depression and anxiety symptoms** and **dependent interpersonal stress** and an increase in **self-esteem** in boys (Rueger 2008, Auerbach 2011). However, a statistically significant change in depression, anxiety or self-esteem in girls (Rueger 2008), or different types of anxiety symptoms (social, separation, harm avoidance and physical symptoms) in presence of classmate support could not be demonstrated (Grills-Taquechel 2010) Evidence is of very low certainty and results cannot be considered precise due to limited sample size and lack of data.
Reference(s)	**Articles** Arora PG, Wheeler LA, Fisher S, Barnes J. *A Prospective Examination of Anxiety as a Predictor of Depressive Symptoms Among Asian American Early Adolescent Youth: The Role of Parent, Peer, and Teacher Support and School Engagement*. Cultural Diversity and Ethnic Minority Psychology 2017, 23(4):541–550 Auerbach RP, Bigda-Peyton JS, Eberhart NK, Webb CA, Ho MH. *Conceptualizing the prospective relationship between social support, stress, and depressive symptoms among adolescents*. J Abnorm Child Psychol 2011, 39(4):475-87 Branje SJT, van Lieshout CFM, van Aken MAG and Haselager GJT. *Perceived support in sibling relationships and adolescent adjustment*. Journal of Child Psychology and Psychiatry 2004, 45(8):1385–1396 Branstetter SA, Low S and Furman W. *The influence of parents and friends on adolescent substance use: A multidimensional approach*. J Subst Use 2011, 16(2):150–160

Burke T, Sticca F, Perren S. *Everything's Gonna be Alright! The Longitudinal Interplay among Social Support, Peer Victimization, and Depressive Symptoms*. J Youth Adolescence 2017, 46:1999–2014

Burton E, Stice E and Seeley JR. *A prospective test of the stress-buffering model of depression in adolescent girls: No support once again*. Journal of Consulting and Clinical Psychology 2004, 72(4):689-697

Cavanaugh AM and Buehler C. *Adolescent loneliness and social anxiety: The role of multiple sources of support*. Journal of Social and Personal Relationships 2016, 33(2):149–170

Colarossi LG and Eccles JS. Differential effects of support providers on adolescents' mental health. Social Work Research 2003, 27(1):19-30

Cotter KL, Wu Q and Smokowski PR. *Longitudinal Risk and Protective Factors Associated with Internalizing and Externalizing Symptoms Among Male and Female Adolescents*. Child Psychiatry Hum Dev 2016, 47(3):472-85

Cotter KL and Smokowski PR. *An Investigation of Relational Risk and Promotive Factors Associated with Adolescent Female Aggression*. Child Psychiatry Hum Dev 2017, 48(5):754-767

Dauber SE, Paulson JF and J. A. Leiferman. *Race-specific transition patterns among alcohol use classes in adolescent girls*. J Adolesc 2011, 34(3):407–420.

Fanti KA, Demetriou AG and Hawa VV. *A longitudinal study of cyberbullying: Examining risk and protective factors*. European Journal of Developmental Psychology 2012, 9(2):168-181

Gagné A-S, Blackburn M-È, Auclair J, Brault M-C, Dion J. *Appearance Esteem Trajectory According to Three Different Sources of Support Among Adolescents Over a School Year*. J Youth Adolescence 2020, 49:2190–2202.

Grills-Taquechel AE, Norton P and Ollendick TH. *A longitudinal examination of factors predicting anxiety during the transition to middle school*. Anxiety Stress Coping 2010, 23(5): 493–513

Jaycox LH, Stein BD, Paddock S, Miles JN, Chandra A, Meredith LS, Tanielian T, Hickey S and Burnam MA. *Impact of teen depression on academic, social, and physical functioning*. Pediatrics 2009, 124(4):e596-e605

Jiang X, Huebner ES, Siddall J. *A Short-Term Longitudinal Study of Differential Sources of School-Related Social Support and Adolescents' School Satisfaction*. Soc Indic Res 2013, 114:1073–1086

Khatib Y, Bhui K and Stansfeld SA. *Does social support protect against depression & psychological distress? Findings from the RELACHS study of East London adolescents*. Journal of Adolescence 2013, 36:393–402

Kendrick K, Jutengren G and Stattin H. *The protective role of supportive friends against bullying perpetration and victimization*. Journal of Adolescence 2012, 35:1069–1080

Kirsch G, McVey G, Tweed S, Katzman DK. *Psychosocial Profiles of Young Adolescent Females Seeking Treatment for an Eating Disorder*. Journal of Adolescent Health 2007, 40:351–356

Klima T and Repetti RL. *Children's Peer Relations and Their Psychological Adjustment Differences between Close Friendships and the Larger Peer Group*. Merrill-Palmer Quarterly 2008, 54(2):151-178

Lester L, Cross D, Shaw T & Dooley J. Adolescent bully-victims: Social health and the transition to secondary school. Cambridge Journal of Education 2012, 42(2):213-233

Lester L and Cross D. *The Relationship Between School Climate and Mental and Emotional Wellbeing Over the Transition from Primary to Secondary School*. Psych Well-Being 2015, 5:9

Magro SW, Utesch T, Dreiskämper D, Wagner J. *Self-esteem development in middle childhood: Support for sociometer theory*. International Journal of Behavioral Development 2019, 43(2):118–127

Nilsen W, Karevold E, Røysamb E, Gustavson K, Mathiesen KS. *Social skills and depressive symptoms across adolescence: Social support as a mediator in girls versus boys*. Journal of Adolescence 2013, 36:11–20

Pisarska A and Ostaszewski K. *Factors associated with youth gambling: longitudinal study among high school students*. Public Health 2020, 184:33-40

Pössel P, Burton SM, Cauley B, Sawyer MG, Spence SH and Sheffield J. *Associations between Social Support from Family, Friends, and Teachers and depressive Symptoms in Adolescents*. J Youth Adolescence 2018 47(2):398-412

Rawana KS. *The relative importance of body change strategies, weight perception, perceived social support, and self-esteem on adolescent depressive symptoms: Longitudinal findings from a national sample*. Journal of Psychosomatic Research 2013, 75:49–54

Rosario M, Salzinger S, Feldman RS, Ng-Mak DS. *Intervening processes between youths' exposure to community violence and internalizing symptoms over time: The roles of social support and coping*. Am J Community Psychol 2008, 41:43–62

Rueger SY, Malecki CK and Demaray MK. *Relationship Between Multiple Sources of Perceived Social Support and Psychological and Academic Adjustment in Early Adolescence: Comparisons Across Gender*. J Youth Adolescence 2010, 39:47–61

Slomkowski C, Rende R, Conger KJ, Simons RL, Conger RD. *Sisters, Borthers, and Delinquency: Evaluating Social Influence during Early and Middle Adolescence*. Child Development 2001, 72(1):271–283

Smokowski PR, Guo S, Cotter KL, Evans CBR, Rose RA. *Multi-level risk factors and developmental assets associated with aggressive behavior in disadvantaged adolescents*. Aggressive Behavior 2016, 42:222–238

Smokowski PR, Guo S, Cotter KL, Evans CBR, Bacallao M, Cotter KL. *Risk and protective factors across multiple microsystems associated with internalizing symptoms and aggressive behavior in rural adolescents: Modeling longitudinal trajectories from the Rural Adaptation Project*. American Journal of Orthopsychiatry 2017, 87(1):94–108

Stice E, Ragan J and Randall P. *Prospective Relations between Social Support and Depression: Differential Direction of Effects for Parent and Peer Support?* Journal of Abnormal Psychology 2004, 113(1):155–159

Swirsky JM and Xie H. *Peer-Related Factors as Moderators between Overt and Social Victimization and Adjustment Outcomes in Early Adolescence*. Journal of Youth and Adolescence 2021, 50:286–297

Way N and Robinson MG. *A longitudinal study of the effects of family, friends, and school experiences on the psychological adjustment of ethnic minority, low-SES adolescents*. Am J Community Psychol 2007, 40:194–213

Wentzel KR, Tomback R, Williams A and McNeish D. *Perceptions of competence, control, and belongingness over the transition to high school: A mixed-method study*. Contemporary Educational Psychology 2019, 56:55–66

Wright MF and Wachs S. *Does social support moderate the relationship between racial discrimination and aggression among Latinx adolescents? A longitudinal study*. Journal of Adolescence 2019, 73:85-94

Young JF, Berenson K, Cohen P and Garcia J. *The role of parent and peer support in predicting adolescent depression: A longitudinal community study*. Journal Of Research On Adolescence 2005, 15(4):407–423

Väänänen JM, Marttunen M, Helminen M and Kaltiala-Heino R. *Low perceived social support predicts later depression but not social phobia in middle adolescence*. Health Psychology and Behavioral Medicine 2014, 2(1):1023-1037

Zimmerman MA, Ramirez-Valles J, Zapert KM and Maton KI. *A longitudinal study of stress-buffering effects for urban African-American male adolescent problem behaviors and mental health*. Journal of community psychology 2000, 28(1):17–33

Measurement instruments for risk factors

Risk factor	Author, year	Name of instrument	Content of the instrument
Multidimensional Scale of Perceived Social Support (MSPSS)			
Supportive social relations	Fanti, 2012	Multidimensional Scale of Perceived Social Support (MSPSS)	This instrument was used to measure supportive relationships within three contexts: family (e.g. "I get the emotional support I need from my family"), friend (e.g. "I can count on my friend when things go wrong"), and school (e.g., "The staff at my school provides me the support and encouragement that I need"). The participants respondent on a 4-point scale (from 0=Not at all true to 3=Definitely true). Prior work has provided evidence that the MSPSS is a valid and reliable measure of perceived social support during adolescence. See Zimet 1988 for more information.
Perceived friends' support	Gagné, 2020		Examples of these items are "I can count on my friends when things go wrong"

			and "I can talk about my problems with my friends." Higher scores indicated greater support from friends. See Zimet 1988 for more information.
Social support from friends	Khatib, 2013		The MSPSS is a 12-item self-report measure, where each statement is scored on a 7-point Likert scale and assesses social support from three sources: family, **friends** and significant other or special person. There was good internal consistency on the MSPSS at baseline (Cronbach's alpha 0.90) and for each ethnic group (White UK 0.91; Asian Bangladeshi 0.90; Black 0.87). See Zimet 1988 for more information.
	Pössel, 2018		Students are asked to rate how strongly they agree or disagree with each statement on a 7-point Likert type scale (1 = very strongly disagree to 7 = very strongly agree). For the present study only the family and peer items were used including 4 items measuring family support (e.g. "My family really tries to help me.") and 4 items measuring friend support (e.g. "I can count on my friends when things go wrong."). See Zimet 1988 for more information.
Perceptions of Peer Social Support Scale			
Peer support	Lester, 2012 Lester, 2015	Peer support at school scale, adapted from the 24-item Perceptions of Peer Social Support Scale	11 items (how often would other students: choose you on their team at school; tell you you're good at things; explain something if you didn't understand; invite you to do things with them; help you if you are hurt; miss you if you weren't at school; help you if something is bothering you; ask to work with you on group work; help you if other students are treating you badly; ask you to join in when you are alone; and share their things with you?) were measured on a three-point scale (1 = never, 2 = sometimes, 3 = lots of times). A peer support score at each time-point was calculated for each student by averaging all items, with a higher score reflecting greater feelings of peer support.
School Success Profile (SSP)			
Friend Support	Cotter, 2016 Cotter, 2017 Smokowski, 2016 Smokowski, 2017	School Success Profile (SSP)	Friend Support was measured with a 5-item scale. Example items included: "I can count on my friends for support" and "I can trust my friends." Each item was rated on a 3-point Likert Scale (Not Like Me, A Little Like Me, or A Lot Like Me). See Bowen 2008 for more information.
Child and Adolescent Social Support Scale (CASSS)			
Social support from friends	Wright, 2019	Child and Adolescent Social Support Scale (CASSS)	See Malecki 2001 for more information.
Friend support and classmate support	Rueger, 2008	Child and Adolescent Social Support Scale (CASSS)	The CASSS is a 60-item, self-report measure of perceived social support. Students rate the frequency with which

			they perceive supportive behaviors from parents, teachers, **classmates, close friends**, and the school (from 1-never to 6-very often). There are 12 supportive behaviors for each source of support. Furthermore, among those 12 behaviors for each source, 3 assess emotional support (i.e., feeling loved or cared for), 3 assess informational support (i.e., receiving advice or information), 3 assess instrumental support (i.e., time, resources, financial support), and 3 assess appraisal support (i.e., feedback). In addition to rating the frequency with which they perceive these support behaviors, students also rate the importance of those behaviors to them (from 1-not important to 3-very important); *however, only frequency ratings were used in the current study*. See Malecki 2001 for more information.
Network of Relationships Inventory (NRI)			
Perceived social support	Stice, 2004	Network of Relationships Inventory (NRI)	See Network of Relationships Inventory.doc
Friendship support	Branstetter, 2011	15 items of the Network of Relationships Inventory (NRI)	The support factor was comprised of five 3-item scales: 1) seeking a safe haven, 2) providing a safe haven, 3) seeking a secure base, 4) providing a secure base, and 5) companionship. The support scales used in the present study were designed to assess aspects of attachment, caregiving, and affiliation. See Network of Relationships Inventory.doc
Peer support	Burton, 2004	12 items from the Network of Relationships Inventory (NRI)	Items assessed companionship, guidance, intimacy, affection, admiration, and reliable alliance from and peers. See Network of Relationships Inventory.doc
Inventory of Parent and Peer Attachment (IPPA)			
Friendship support	Burke, 2017	Selected items from the Inventory of Parent and Peer Attachment (IPPA)	Three items were selected for friendship support: "My friends accept me as I am," "I trust my friends," "My friends understand me." Participants rated each item on a Likert scale from 1 (not true) to 4 (true).
Social support (from peers)	Rosario 2008	Inventory of Parent and Peer Attachment (IPPA)	A set of 25 items, each on a 5-point Likert scale, were used to ask about the youth's close friends (e.g. "I like to get my friends' opinion on things I'm concerned about " or "My friends listen to what I have to say"). See Inventory of parent and peer attachment.doc.
Perceived Social Support from Friends (PSS-Fr)			
Friend support	Pisarska, 2020	Perceived Social Support from Friends (PSS-Fr).	See Procidano 1983 for more information.
	Zimmerman, 2000	Shortened version of Procidano (1983)	The shortened friend-support scale included 10 items, with a possible range from 10 to 50.

		Parents and Friends Scale	See Procidano 1983 for more information.
Friendship support	Way, 2003	Perceived Social Support Scale for Friends (PSS FR).	Students were asked to respond yes, no, or don't know to 20 items concerning their experiences with their friends. In the present sample, this measure demonstrated good internal consistency. For the purposes of the present analyses, the total number of positive responses was used as a summary score of the perceived quality of friendship support. See Procidano 1983 for more information.
Social Support Scale for Children (SSSC)			
Friend support and classmate support	Grills-Taquechel, 2010	Social Support Scale for Children (SSSC)	24 antithetical statements (e.g., "Some kids have a close friend who they can tell problems to BUT other kids don't have a close friend who they can tell problems to") that children choose and rate as really true or sort-of-true for themselves. This scale yields four subscales corresponding to individuals from whom support is received, including **classmates, and close friends**. See Harter 2012 for more information.
Close friend support	Klima, 2008	Close Friend Support subscale of the Social Support Scale for Children (SSSC)	The scale assesses whether children perceive that they have a caring, understanding friend to whom they can disclose problems and feelings. Questions are presented in a structured alternative format designed with a 4-point response scale (see above). Responses to six questions are averaged, with higher scores indicating more support. The Close Friend Support subscale has been shown to be valid and reliable. The internal consistency of the scale was acceptable to good. See Harter 2012 for more information.
Peer and classmate social support	Auerbach, 2011	The Social Support Scale for Children and Adolescents (SSSCA)	The SSSCA is an 18-item self-report questionnaire. Each item requires the participant to first choose between two statements that best reflects the type of person they are more like. Then, the participant is asked to determine whether the chosen statement is "really true for me" or "sort of true for me." Item scores range from 0–3, and higher scores are indicative of a higher perceived quality of social support. Possible total scores range from 0–54, and for each subscale, the possible range is from 0 to 18. The SSSCA assesses perceived social support in the domains of **peer, parent, and classmate relationships** with lower scores representing less perceived social support. Peer support describes general non-specific friendship and questions included, "Some kids have a close friend who they can tell problems to" versus

			"Other kids don't have a close friend who they can tell problems to." In contrast, classmate support examines relationships as it relates to general support as well as peer victimization in the classroom and/or school. Questions include (a) "Some kids have classmates they can become friends with" versus "Other kids don't have classmates that they can become friends with" and (b) "Some kids have classmates who sometimes make fun of them" versus "Other kids don't have classmates who make fun of them." See Harter 2012 for more information.
Other scales			
Peer support	Cavanaugh, 2015	- 5-item friend support (Richman 1997) - 7-item children's friendship scale (Berndt 1986)	- A sample item was "I can trust my friends." The response format ranged from 0 (not like me) to 2 (a lot like me). Items were averaged and higher scores indicated greater support from close peers. - A sample item was "when you do a good job on something how often does this friend praise and congratulate you?" The response format ranged from 1 (never) to 5 (every day). Items were averaged, and higher scores indicated greater support.
	Swirsky 2021	0-4 Likert scale	Participants reported the amount of peer support they received via a 0 (none) – 4 (all the time) Likert scale. Peer support was measured as the mean across three items in the Spring semester (Time 2): How often has someone: "given you help when you needed it", "comforted you when you were upset or lonely", and "stood up for you".
	Wentzel 2019	Peer Social Support subscale of the Classroom Life Measure	Sample items are: "My classmates really care about me." Responses were made on a five-point scale, 1=Never, 2=Seldom, 3=Sometimes, 4=Often, and 5=Always. Items were averaged to form scale scores that were then standardized.
	Jaycox, 2012	Social Adjustment Scale for Youth	See Weissman 1980 for more info (Social Adjustment Scale – Self-Report which was modified for children).
	Colarossi, 2013	Iowa Youth and Families Inventory	Six-items for peer support were measured on a 5-point Likert scale from 1 = low support to 5 = high support. Examples of scale items include 'In the last month, how often did your friends … help you with something that was important to you? … give you advice? … let you know she/he really cares about you? The internal consistency of the scale was good. See Harter 2012 for more information. See Conger (1986) for more information (little info on questions available).

	Dauber, 2011	Researcher-made questionnaire	Adolescents named up to five male and five female friends, and indicated for each whether they had gone to the friend's house, hung out after school, talked on the phone, spent time together on the weekend, and discussed a problem. An index of peer support was created by summing responses for each friend and averaging across all friends, yielding one peer support score. Adolescents who did not list any friends were given a score of zero. Total scores ranged from 0 to 5 with higher scores indicating more peer support.
Support from peers	Arora, 2017	Researcher-made questionnaire?	Adolescents reported on support from peers (i.e., "My friends help me when I'm having a hard time," "My friends are there when I need them," "I feel that I can talk to my friends about my problems"; 5 items). Adolescents self-reported on a 4-point Likert scale (1 = strongly disagree, 4 = strongly agree).
Supportive friends	Kendrick 2012	Friendship Quality Questionnaire	See Parker&Asher_1993_friendship quality scale. The items that were adapted from this scale were: 'My VIP stands by me if others talk about me behind my back'; 'My VIP would like me even if nobody else did'; 'My VIP says "I'm sorry" when he or she has hurt my feelings or been mean'; 'My VIP doesn't tell my secrets to others'; and 'My VIP pays attention to my feelings'. Two additional items were included: 'My VIP supports me when I've had an argument with my parents/teachers'; and 'My VIP keeps his or her promises'. Possible responses were on a five-point scale, ranging from: (1) 'Don't agree at all', to (5) 'Agree completely'.
Peer Support for Learning	Jiang 2012	Peer Support for Learning subscale of the Student Engagement Inventory (SEI)	Full SEI: see document Student Engagement Inventory.pdf Subscale consists of Six questions designed to measure the extent to which each student perceives their learning is supported by their peers (e.g., "Other students at school care about me").
Conditional support from peers	Kirsch 2007	Scale designed by Marold (doctoral thesis).	Items tap the extent to which support from mothers and fathers is conditional upon their children meeting high parental expectations. The peer scale measured the extent to which support was perceived as conditional upon meeting peer expectations that involved standards of appearance, behavior, and attitudes.
Anticipated peer support	Young 2005	6 interview questions that were answered "yes" or "no".	Adolescents responded to 4 questions about the availability of support from friends. These items were: "Do you have one or more friends who... (1) you can talk to about almost anything, (2) will stick up for you no matter what, (3) would turn to you for advice or help, (4) really understands you?'' Adolescents

			were also asked: "Have you had a friend for at least 1 year?" and "Do you make friends easily?" Because each of these items had a fairly low frequency of "no" responses, they were combined to form one dichotomous variable. An adolescent was coded as having low anticipated support from peers if s/he answered no to any of these six items. The remaining adolescents were coded as having high anticipated support from peers.
Perceived social support	Magro 2019	Peer and family climate subscales of the German and Dutch versions of the Kid-KINDL-R	Peer support was measured with three items reflecting peer behaviors over the last week ("I got along well with my friends," "I played with friends," and "Other kids liked me") on a 5-point scale.
Social support	Väänänen 2014	Perceived Social Support Scale-Revised	The PSSS-R measures people's subjective perceptions of social support and emotional closeness. It contains 12 items on a 5-point Likert-type scale. Factor-analytically derived sum scores were used for addressing perceived support from family, friends, and significant others (each ranging 4–20). High sum scores indicate high perceived social support. PSSS-R sum scores were used as a continuous variable.
Warmth-support between siblings	Slomkoswki, 2001	8-item scale	Each child reported, using a 7-point response format 1=never behaved that way to 7=always behaved that way how often his or her sibling behaved in certain ways during the previous month. The 8 items included behaviors such as: listens carefully to your point of view, acts loving and affectionate, lets you know he or she appreciates you and your ideas. The items were summed to create an indicator of "warmth–support"
Perceived sibling support	Branje, 2004	Relational Support Inventory (RSI)	The inventory involves 24 questions representing four dimensions of perceived support measured by six items each along a 5-point Likert scale ranging from very untrue of this person (1) to sometimes untrue, sometimes true of this person (3) to very true of this person (5). The first support dimension, perceived Quality of Information, assesses the quality of information and withholding of information. A sample item is: 'This person explains or shows how I can make or do something.' The second support dimension is perceived Respect for Autonomy and assesses respect for autonomy and limit setting. For example, 'This person lets me solve problems as much as possible on my own but also provides help when I ask for it.' The third support dimension is perceived Emotional Support and assesses warmth as opposed to hostility. A sample item is: 'In this person's view,

			I can't do anything right: he/she is always criticizing me.' The fourth support dimension is perceived Convergence of Goals and assesses the perceived level of convergence as opposed to divergence of goals. For example: 'This person and I have many conflicts with regard to my school achievement, future, or career opportunities' (reverse coded). Siblings judged the support they perceived from each other and from their father, mother and best friend. RSI total scores were averaged across all 24 items.

Thematic category: Social skills

Characteristics of included studies

Author, year, Country	Study design	Population	Risk factor [information on measurement instrument in table below conclusions]	Remarks
Le Grange , 2017, Australia	Observational: prospective cohort study (3 timepoints in a 5-year period, part of a larger study outside our age range)	1300 adolescents, aged 11-16 years (mean age unknown). A representative sample of 2,443 infants was recruited through selected Maternal and Child Health Centers across both urban and rural areas in the State of Victoria.	Social skills (Social Skills Rating System, self-reported) (measured at 11-12 years (T1), and 13-14 years (T2))	Outcomes measured: At 15-16 years (T3): Abnormal eating attitudes and behaviour, based on the Drive for Thinness and Bulimia Subscales of the Eating Disorder Inventory (EDI), and an adapted version of the EDI Body Dissatisfaction Subscale suitable for administration in a (non-clinical) general population sample Statistics: construction of a path model

Synthesis of findings

Outcome	Risk factor	Effect Size	#studies, # participants	Reference
Abnormal eating attitudes and behaviour at T3	Social skills at T1	Not statistically significant: £† (p>0.05)	1, 1300	Le Grange, 2017
	Social skills at T2	Not statistically significant: £† (p>0.05)		

£ No raw data available

† Imprecision (lack of data)
Study limitations

Author, Year	Inappropriate eligibility criteria	Inappropriate methods for exposure and outcome variables	Not controlled for confounding	Incomplete or inadequate follow-up	Other limitations
Le Grange, 2017	Unclear, not clear how the Maternal and Child Health centres were selected and how the sample of children was recruited from these centres.	Yes, exposure and outcome measures were self-reported.	No, all other individual, interpersonal and family factors are taken into account in the path analysis (but not controlled for baseline abnormal eating, see also other limitations).	No, the study was a 15-wave study over 28 years, 28% of the youth were still enrolled at the last wave. Although proportionately more families from a lower SES background, or parents who were not born in Australia, have been lost to the study, there were no significant differences on any child characteristics assessed in infancy between the retained cohort at 15–16 years and those no longer participating.	We were not able to control for earlier Abnormal Eating Attitudes and Behaviors and hence, although the predictor variables were all antecedent in time to the outcome, it is possible that they could have arisen as consequences of previous Abnormal Eating Attitudes and Behaviors. However, the authors think this is unlikely.

Certainty of the body of evidence

Social skills	Initial grading Low [C]	Downgrading due to
Limitations of study design	-1	See table 'Study limitations'
Imprecision	-1	Lack of data
Inconsistency	0	
Indirectness	0	
Publication bias	0	
		Upgrading due to
Large magnitude of effect	0	
Dose-response gradient	0	
Plausible confounding	0	
CERTAINTY (GRADE)	**Final grading Very low [D]**	

Conclusion	There is limited evidence concerning the risk of abnormal eating attitudes and behaviour in case of having social skills. A statistically significant decreased risk of abnormal eating attitudes and behaviour in case of having social skills could not be demonstrated (Le Grange 2017). Evidence is of very low certainty and results cannot be considered precise due to lack of data.
Reference(s)	**Articles** Le Grange D, O'Connor M, Hughes EK, Macdonald J, Little K, Olsson CA. *Developmental antecedents of abnormal eating attitudes and behaviors in adolescence*. Int J Eat Disord 2014, 47(7):813-24

Measurement instruments for risk factors

Risk factor	Name of instrument	Content of the instrument
Social skills	Social Skills Rating System	The Social Skills Rating System allows to obtain a more complete picture of social behaviors from teachers, parents, and students themselves. The scale evaluates a broad range of socially validated behaviors-behaviors that affect teacher-student relationships, peer acceptance, academic performance, and more. The Social Skills Scale measures positive social behaviors: Cooperation, Empathy, Assertion, Self-Control, Responsibility. Example item: "makes friends easily" (checklist only available after payment)