

Report prepared for Brå by David P. Farrington,
Friedrich Lösel, Maria M. Ttofi and Nikos Theodorakis

School Bullying, Depression and Offending Behaviour Later in Life

An Updated Systematic Review of Longitudinal Studies

brå

National Council for Crime Prevention

School Bullying, Depression and Offending Behaviour Later in Life

An Updated Systematic Review
of Longitudinal Studies

Brå – a centre of knowledge on crime and measures to combat crime

The Swedish National Council for Crime Prevention (Brottsförebyggande rådet – Brå) works to reduce crime and improve levels of safety in society by producing data and disseminating knowledge on crime and crime prevention work and the justice system's responses to crime.

Production:

Brottsförebyggande rådet/The Swedish National Council for Crime Prevention
Box 1386, SE-111 93 Stockholm, Sweden
Phone +46 (0)8-401 87 00, fax +46 (0)8-411 90 75, e-mail info@bra.se, www.bra.se

Authors: David P. Farrington, Friedrich Lösel, Maria M. Tfofi, Nikos Theodorakis

Cover Illustration: Helena Halvarsson

Printing: Edita Västerås 2012

© Brottsförebyggande rådet 2012

ISSN 1100-6676

ISBN 978-91-86027-89-6

URN:NBN:SE:BRA-470

Contents

Foreword.....	5
Executive Summary.....	7
Acknowledgements.....	10
1. Introduction.....	12
1.1 Background.....	12
1.2 Objectives of the Review and Main Questions Addressed.....	14
2. Methods.....	16
2.1 Criteria for Inclusion or Exclusion of Studies.....	16
2.2 Searching Strategies.....	21
2.3 Screening of Reports.....	25
2.4 Included and Excluded Studies.....	28
2.5 Combining Effect Sizes within a Report: Bullying Perpetration/ Victimization versus Offending.....	44
2.6 Combining Effect Sizes within a Report: Bullying Perpetration/ Victimization versus Depression.....	50
2.7 Combining effect sizes across reports relating to the same longitudinal study.....	59
3. Bullying Perpetration versus Offending.....	60
3.1 Unadjusted and Adjusted Effect Sizes.....	60
3.2 Moderator Analyses.....	62
3.3 Publication Bias Analyses.....	64
4. Bullying Victimization versus Depression.....	66
4.1 Unadjusted and Adjusted Effect Sizes.....	66
4.2 Moderator Analyses.....	68
4.3 Publication Bias Analyses.....	69
5. Bullying Victimization versus Offending.....	71
5.1 Unadjusted and Adjusted Effect Sizes.....	71
5.2 Moderator Analyses.....	73
5.3 Publication Bias Analyses.....	74
6. Bullying Perpetration versus Depression.....	75
6.1 Unadjusted and Adjusted Effect Sizes.....	75
6.2 Moderator Analyses.....	75
6.3 Publication Bias Analyses.....	77
7. Conclusions and Recommendations.....	79
References.....	83
Appendices.....	95
Appendix 1: Combining Data on Offending Outcomes.....	95
Appendix 2: Combining Data on Depression Outcomes.....	100

Foreword

Bullying is a problem among children all over the world. In an earlier report in this series, two of the authors of this study have shown that systematic school programs have proven to be effective in preventing bullying. This was an encouraging result. A further question of interest is that of whether bullying also influences the bullies and the victims later on in life in terms of subsequent offending and mental health problems. The answer to this question would reveal whether the prevention of bullying also constitutes a means of preventing future crime and mental health related issues. This is the question answered by the four authors of this report on the basis of a systematic review that includes a number of statistical meta-analyses.

There are never sufficient resources to conduct rigorous evaluations of all the crime prevention measures employed in an individual country such as Sweden. Nor are there resources to conduct scientific studies of all of the effects produced by e.g. early risk-factors on later offending. For these reasons, the Swedish National Council for Crime Prevention (Brå) has commissioned distinguished researchers to conduct a series of international reviews of the research published in these fields.

This report presents a systematic review, including a number of statistical meta-analyses, of the impact of bullying on later offending and depression, with regard to both the bullies and those exposed to bullying. The work has been conducted by Professor David P. Farrington, Professor Friedrich Lösel, Dr. Maria M. Tfofi, and Ph.D. candidate Nikos Theodorakis, all of Cambridge University.

The study follows the rigorous methodological requirements of a systematic review. The analysis combines the results from a substantial number of studies that are considered to satisfy a list of empirical criteria for measuring the correlations of bullying perpetration and victimization with offending and depression as reliably as possible. The meta-analysis then uses the results from these

previous studies to calculate and produce a robust overview of the impact of bullying on negative outcomes later in life.

The systematic review, and the statistical meta-analyses, in this case builds upon a large number of scientific studies from different part of the world, producing highly relevant findings on the impact of bullying among children on offending and depression later in life. Although some important questions remain unanswered, the study provides the most accessible and far-reaching overview of this important issue that has been produced to date.

Stockholm, June 2012

Erik Wennerström
Director-General

Executive Summary

School bullying is a serious problem in many countries. Most research on this topic concentrates on the prevalence, origins and prevention of bullying and victimization (being bullied). However, there is also an increasing body of knowledge about the long-term negative impact of school bullying and victimization on later life outcomes. This report presents results from a comprehensive systematic review on the extent to which school bullying and victimization predict offending and depression later in life.

The results are mainly based on prospective longitudinal studies and arise from the activities of a two-year international research network. Principal investigators and researchers of 29 longitudinal studies participated in this network, providing published and unpublished data for our project on *Health and Criminal Outcomes of Children involved in School Bullying* that is carried out for the Campbell Collaboration Crime and Justice Group. Results from retrospective longitudinal studies, found in the published literature, are also included for depression as the outcome measure.

Two predictors (bullying perpetration and victimization), measured in the school years, and two outcomes (offending and depression), measured in later life, were studied in four meta-analyses. These meta-analyses specify the strength of the relationship of school bullying and victimization with later offending and depression. Further analyses followed methodological strategies of the *Cambridge Quality Checklist on Risk Factor Research* and investigated whether effects remain significant after controlling for other major childhood risk factors, which were significantly related to both the predictors and the outcomes. Such risk factors varied across the primary studies and covered a broad range of individual, family, neighbourhood and other variables.

Our results are based on extensive searches of the literature. Electronic databases and journals were searched from the inception of each database or journal up to the end of December 2011. In total, we have searched 63 journals and 19 databases. Explicit criteria for

inclusion or exclusion of studies in our meta-analyses were set in advance. In total, we located 661 reports that addressed the association of school bullying with internalizing problems (e.g. anxiety, depression, self-esteem, etc.) and externalizing problems (e.g. aggressive behaviour, conduct problems, offending, etc.). All reports were screened in line with our inclusion and exclusion criteria and classified in five different categories.

Further to a detailed screening of all manuscripts, 48 reports from 29 longitudinal studies were included in our systematic review on the association of bullying perpetration and victimization with offending later in life; and 75 reports from 49 longitudinal studies were included in our systematic review on the association of bullying perpetration and victimization with depression later in life. Not all studies provided effect size data for our meta-analyses. Clear rules were set in advance for combining effect sizes within a report as well as for combining effect sizes across reports relating to the same longitudinal study.

For all 48 reports on offending and 75 reports on depression, detailed features of the studies were coded such as: the age at which school bullying was measured; the age at which outcome measures were reported; the length of follow-up period; and the number of covariates (i.e. other major childhood risk factors) controlled for in the school years. These features were later included in various moderator analyses in an attempt to explain variations in effect sizes across studies.

As expected, bullying perpetration at school was a highly significant predictor of offending on average six years later in life. The summary Odds Ratio (OR) of the unadjusted effect size across 18 studies was $OR = 2.64$. After controlling for other childhood risk factors, the adjusted effect size across 15 studies was $OR = 1.89$ and still significant. This value of the OR suggests that being a bully increases the risk of later becoming an offender by more than half.

The probability of being depressed an average of seven years later in life was significantly greater for victims of school bullying than for other students. The unadjusted effect size across 30 studies was $OR = 2.05$ and the adjusted effect size (after controlling for childhood risk factors) across 19 studies was $OR = 1.71$. This value of the OR suggests that being a victim of bullying increases the risk of later becoming depressed by about half.

Bullying victimization was a weaker predictor of offending. The unadjusted effect size of $OR = 1.40$ across 14 studies was statistically significant. The adjusted effect size of $OR = 1.14$ across 12 studies was nearly significant. This value of the OR suggests that being a victim of bullying increases the risk of later becoming an offender by only 10%.

Bullying perpetration was significantly related to later depression. The unadjusted effect size across 16 studies was $OR = 1.61$.

The adjusted effect size across 13 studies was smaller (OR = 1.41), but still statistically significant. This value of the OR suggests that being a bully increases the risk of later becoming depressed by about 30%.

Some moderator analyses showed that the effect sizes were smaller when the outcomes were measured at older ages (i.e. with a longer time interval since the predictors were measured) and when more childhood risk factors were controlled for. There was no evidence of publication bias in any of our analyses.

This report provides the most detailed, comprehensive and up-to-date scientific evidence on the detrimental effect of school bullying and victimization on children's mental health and psychosocial development later in life. Our findings clearly show that bullying and victimization significantly predict later offending and depression, even after controlling for other major childhood risk factors. Therefore, children involved in school bullying as perpetrators or victims are high-risk youth, and it is concluded that effective high quality anti-bullying programmes are essential. Our previous systematic reviews of such programmes show that many are effective. In light of the results of the present report, these programmes could be viewed as an early form of preventing crime as well as a method of promoting health. Therefore, our research findings have important implications for policy and practice. They underline the need for school communities and relevant authorities to create a violence-free school environment and the need to devise and implement measures to interrupt the continuity from school bullying to later adverse life outcomes.

Acknowledgements

This project synthesizes data from published and unpublished studies. We are most grateful to all principal investigators of prospective longitudinal studies and to their collaborators for conducting special analyses for the aims of this project.

The following scholars have conducted special analyses for our review: Robert D. Abbott, Louise Arseneault, Leena K. Augimeri, Margit Averdijk, Silvia Azzouzi, Doris Bender, Catrien Bijleveld, Lyndal Bond, William Bor, Lucy Bowes, Deborah Capaldi, Richard F. Catalano, Pernille Due, Ben Edwards, Leonard Eron, Manuel Eisner, Dorothy L. Espelage, David M. Fergusson, Victor van der Geest, Sheree J. Gibb, Nancy Guerra, Kevin P. Haggerty, HonaLee Harrington, Sheryl A. Hemphill, Jan Hendriks, David Henry, Todd I. Herrenkohl, L. John Horwood, Renate Houts, L. Rowell Huesmann, Caroline Hunt, Atsushi Igarashi, Hiromitsu Inoue, Depeng Jiang, Kristin Kendrick, Martin Killias, Min Jung Kim, Birgitta Kimber, Aneta Kotevski, Tara R. McGee, John J. McGrath, Susan McVie, Terrie E. Moffitt, Barbara Müller, Jake M. Najman, Katsumi Ninomiya, Yasuyo Nishino, Michael O'Callaghan, Dan Olweus, Metin Özdemir, Jennifer Renda, Denis Ribeaud, Christina Salmivalli, Rolf Sandell, James G. Scott, Rebecca Stallings, Hakan Stattin, Patrick Tolan, John W. Toumbourou, Tatsuo Ujiie, Richard VanAcker, Suzanne Vassallo, Margaret Walsh, Gail M. Williams, Chika Yamamoto.

The following scholars have kindly produced datasets relevant to our review that have not yet been analyzed for this study: Patricia Cohen (for the New York State Longitudinal Study), Karl Hill (for the Seattle Social Development Project), and Antonio Castro Fonseca (for the Portuguese Longitudinal Study).

Some studies would have been includable if further information could have been obtained. These include intervention studies (with before and after measures) and prospective longitudinal studies. We would nevertheless like to thank Herbert Marsh, Mühlbacher Moritz, Marius K Nickel, Jessica Robert, Melissa DeRosier, and

Per-Olof Wikstrom for trying to provide some relevant information.

We are also grateful to scholars who have access to longitudinal (or other potentially relevant) datasets and who have confirmed the absence of data for our review: Klaus Boers, Ken Dodge, Margaret E. Ensminger, Vicente Garrido, Britt af Klinteberg, Marc LeBlanc, Alex R. Piquero, Adrian Raine, Carolyn Smith, Terence P. Thornberry, Peter Zettergren, and John Weisz. We would also like to thank scholars who have sent us their published papers on the topic as well as scholars for their helpful comments on the project: Lars Bergman, Frances Gardner, Fred Rivara, and Frank Vitaro.

Finally, we would like to thank Professor David Wilson for his feedback on methodological issues: his continuous support is greatly appreciated. Last but not least, we are most appreciative to Jan Andersson and Erik Grevholm for their continuous support and encouragement in this project.

1. Introduction

1.1 Background

School bullying is a frequent and serious problem in many countries. Scientific interest in this problem and its negative short-term and long-term effects increased after the well-publicized suicides of three Norwegian boys in 1982, which were attributed to the fact that they were severely bullied (Olweus, 1993a). School bullying has gradually become a topic of major public concern via ‘bullying awareness days’, national initiatives in various (European) countries (Smith & Brain, 2000), and anti-bullying research networks across the world (e.g. Anti-Bullying Alliance; BRNET; International Observatory for Violence in Schools; PREVNet).

Recently, school bullying has also attracted a lot of media attention, with articles in major newspapers and magazines reporting cases of children who committed (or attempted) suicide because of severe victimization (being bullied) at school, and parents suing school authorities for their failure to protect their offspring from continued victimization. Examples of articles include: the Daily Mail (UK; September 18, 2009)¹, BBC Online (Wales; April 1, 2010)², and the Boston Globe (USA; December 8, 2010)³. In light of these concerns, it is understandable why school bullying has increasingly become a central topic in intervention and evaluation research (Farrington & Tofi, 2009; Tofi & Farrington, 2011).

But is there indeed scientific evidence about the detrimental effects of school bullying and victimization on children’s physical and mental health? Or could school bullying be viewed as part of a developmental process, ‘one of those school experiences’ that pre-

¹ Webpage: http://article.wn.com/view/2009/09/18/Bullied_girl_15_dies_after_leaping_from_bridge_onto_busy_road/

² Webpage: <http://news.bbc.co.uk/1/hi/wales/8598136.stm>

³ Webpage: http://www.boston.com/news/local/massachusetts/articles/2010/12/14/admission_of_failure/?s_campaign=8315

pare children for the grown-up world, as some sceptics may argue? Of course, schools, like other institutions, will always be places in which the basic human motive of aggression will be seen. However, school bullying should not be confused with more or less normal aggressive interactions such as rough and tumble play. Bullying is a specific form of aggression among children and youth that is not triggered by interpersonal conflicts, but involves an imbalance of power between the perpetrator and the victim and is often rather persistent. Serious bullying is characterized by physical, verbal or psychological attacks and oppression on less powerful youngsters in repeated incidents over a prolonged period of time (Farrington, 1993; Olweus, 1993a).

Bullying is not only an issue of the school climate but it can be highly relevant to the future development of bullying perpetrators and their victims. A number of studies suggest that the prognosis of children who bully and are bullied is not encouraging (Arseneault et al., 2010; Ttofi & Farrington, 2008). When this childhood behaviour is not dealt with, it can spiral out of control in adolescence and adulthood, affecting not only the people themselves but also their relatives and associates. Longitudinal studies have shown that adult violent criminals frequently had school records of bullying and other forms of aggressive behaviour (Luukkonen et al, 2011), suggesting an intra-generational continuity of antisocial or 'externalizing' behaviour. Prospective studies have also pointed out the possibility of inter-generational continuity of school bullying: in the Cambridge Study in Delinquent Development, for example, the bullies at age 14 tended, at age 32, to have children who were also bullies (Farrington, 1993).

For the victims of school bullies, research findings are equally concerning. It has been suggested that victimization (being bullied) may lead to endorsement and establishment of a defeatist self-blaming attitude towards life as well as low self-esteem and depression, causing subsequent problems in the personal, social and work life (Gilmartin, 1987; Matsui et al., 1996; O'Moore et al., 1998; Smith, 1997).

Although the above-mentioned findings suggest that there are longer-term negative outcomes of school bullying and victimization, we do not yet know how strong and robust these relationships are. To date, there has been no attempt to systematically synthesize the results of existing research on the impact of bullying perpetration and victimization on the current physical and mental health of children (based on cross-sectional studies) as well as on their future psychosocial adjustment as adolescents and adults (based on longitudinal studies). In addition, it is not clear whether bullying independently contributes to an undesirable development or whether childhood risk factors cause bullying perpetration and victimization as well as later life outcomes without there being any

causal link between bullying and later life outcomes. For example, it has repeatedly been shown that serious bullying perpetration is not only related to delinquency and violence, but also shares many individual, family, neighbourhood and other risk factors with these behavioural problems (Farrington, 1993; Herrenkohl et al., 2007). Research also suggests that anxiety, depression and social withdrawal may not only be consequences of bullying but also individual characteristics that enhance the risk of being chosen as a victim (Olweus, 1993; Lösel and Bliesener, 2003).

In order to fill the gap in knowledge about the precise quantitative link between school bullying and later life outcomes, the present systematic review has been carried out. We investigated the impact of school bullying perpetration and victimization on later life outcomes, measured using an unbiased standardized effect size, across all available longitudinal studies. In order to reduce the problem of confounded risk factors, we followed a research strategy suggested in the *Cambridge Quality Checklist on Risk Factor Research* (Murray et al., 2009). In addition to bivariate prediction analyses, we undertook meta-analyses of studies that estimated the adjusted effect sizes after controlling for childhood risk factors. The present research concentrates on longitudinal primary studies on the relationship of bullying perpetration and victimization with later outcomes of offending and depression. Studies on other outcomes such as alcohol and drug use, anxiety, self-esteem and violence will be carried out in the future.

1.2 Objectives of the Review and Main Questions Addressed

To date, cross-sectional and longitudinal research has focused on the association of school bullying with internalizing and externalizing behaviour. Many outcomes could be categorized under internalizing syndromes (e.g. anxiety, depression, social withdrawal, and self-worth problems) or externalizing behaviour (e.g. aggression, violence, offending and conduct problems). For the present review, we chose to study one main outcome falling under each category, namely depression (internalizing) and offending (externalizing).

The main objective of our report is two-fold. Firstly, we aim to assess whether bullying at school (perpetration and victimization) is a predictor of depression and offending later in life (unadjusted effect sizes). Secondly, we aim to assess whether these associations are still significant after controlling for other major childhood risk factors (adjusted effect sizes). Our report presents results based on longitudinal studies only. The main questions addressed are as follows:

- Are the victims of school bullying, compared to other children (non-victims or children not involved in bullying), significantly more likely to be depressed later in life? Is the same true of school bullies?
- Are school bullies, compared to other children (non-bullies or children not involved in bullying), significantly more likely to offend later in life? Is the same true of victims of bullying?
- What is the unique contribution of school bullying (perpetration and victimization) to depression and offending later in life? In other words, is school bullying a significant predictor of each outcome after controlling for other potentially relevant childhood risk factors?
- What moderating factors are significantly related to, and might explain, the variability in effect sizes?

We investigate moderators that may explain variability in effect sizes between studies, such as the age at which bullying was measured (Time 1), the age at which the outcome measures were taken (Time 2), the number of covariates controlled for, the length of the follow-up period, the type of longitudinal study (i.e. prospective versus retrospective), and the way in which the outcomes were measured (i.e. official data versus self-reports).

2. Methods

2.1 Criteria for Inclusion or Exclusion of Studies

With regard to eligible study designs, in our review we will only include longitudinal studies:

- using a matched-control design (e.g. based on propensity score matching) to establish whether associations between bullying victimization/perpetration and later adverse outcomes exist independently of possible confounds, such as the ones listed in table 1 entitled ‘List of critical covariates’.
- using statistical controls to establish whether associations between bullying victimization/perpetration and later adverse outcomes exist after controlling for possible confounds such as the ones listed in table 1.

Table 1. List of Critical Covariates

Child covariates
Impulsivity, attention deficits, IQ, school attainment
Parent covariates
Parental antisocial behaviour/ criminality, parental age, parental education, parental mental health, parental substance abuse
Parenting covariates
Low parental supervision, harsh parental discipline, abuse of child, neglect of child, parent-child conflict, inter-parental conflict
Family covariates
Family size, socio-economic status, family income
Wider environmental covariates
Peer delinquency, neighbourhood deprivation, neighbourhood crime, school crime

We will present both unadjusted and adjusted effect sizes summarizing the strength of relationships between predictors and outcomes. Adjusted effect sizes show whether school bullying is followed by a high rate of internalizing/externalizing problems after controlling for earlier risk factors that predict both bullying and the specific outcomes (Murray et al., 2009). We will not examine whether changes in school bullying predict changes in internalizing/externalizing problems, partly because this would require more than two data waves (subsequently excluding studies with only one follow-up) and partly because such change variables are likely to have great variability (Farrington et al., 2011b).

Other criteria for inclusion of reports in the review were as follows:

1. The report clearly indicates that it is concerned with school bullying (perpetration/victimization) and not with other more general forms of aggression among children and youth. We examine school bullying only. The definition of school bullying includes several key elements: physical, verbal, or psychological attack or intimidation that is intended to cause fear, distress, or harm to the victim; an imbalance of power (psychological or physical) with a more powerful child (or children) oppressing less powerful ones; and repeated incidents between the same children over a prolonged period of time (Farrington, 1993; Olweus, 1993a). According to this definition, it is not bullying when two persons of the same strength (physical, psychological or verbal) victimize each other. School bullying can occur in school or on the way to or from school. It is often measured using questionnaires based on the work of Dan Olweus (1993a).
2. A clear measure of depression and/or general offending as an outcome variable is required. Delinquency and violent offending (but not general aggressiveness) can be used as measures of offending if more specific offending items are not available. Our main target is offending outcomes. Therefore, if a study reported outcomes for both offending and violence, we chose to analyze offending. If a study included delinquency and violence, we chose to analyse delinquency. If a study included aggression and violence, we chose to analyse violence. Two reports (Boulton et al., 2010; Smith et al., 2004) provided an effect size of 'behavioural conduct' which might be seen as a proxy for delinquency. However, since it was not clear to what extent 'behavioural conduct' indicated offending, we excluded these studies from the meta-analysis. Another study on gang membership was also excluded (Holmes et al., 1998) since gang membership is not a direct measure of offending. For the meta-analyses relating to depression, the use of anti-depressants was included as an acceptable proxy for depression.

3. The report presents longitudinal data. Subsequently, some papers dealing with depression based on longitudinal studies were excluded because analyses were based on within-wave data, making them essentially cross-sectional in character (e.g. Barbarin, 1999⁴; Grills, 2003⁵, both dealing with depression).
4. Chronologically, the predictor (i.e. bullying perpetration/victimization) precedes the outcome (i.e. depression and offending). Subsequently the Shelley (2009)⁶ study and the Moon *et al* (2011)⁷ study (both dealing with depression) were excluded because of this requirement.
5. We also included follow-up/intervention studies (with before and after measures) since various bullying prevention programmes targeted both health-related problems (such as depression and anxiety) and other behavioural problems. In this case, we sent emails to the evaluators of each programme, asking for specific data analyses for the control group which did not receive the intervention. We did not ask for data analyses based on the experimental children because in the case of efficacious interventions a reduction in bullying might be followed by a reduction in health or other behavioural outcomes. An experimental condition might also have been different from the naturalistic condition in which we were interested. Specifically, we asked the evaluators of the programmes to examine whether bullying at the baseline (i.e. before the implementation of the programme) predicted depression or offending in the follow-up period (i.e. after the implementation of the programme) for the control group only. Other pub-

⁴ The paper shows results within a wave for the Birth-to-Ten (BTT) longitudinal study in South Africa and compares those results with a sample of children who are African American; so, essentially the paper is a cross-national comparison based on cross-sectional data (see Barbarin, 1999: 1351).

⁵ The study involves a 2-year follow up of 77 students from grade six to grade eight. Grills (2003) provides results on the association of peer victimization (Peer Victimization Scale; Neary and Joseph, 1994) but not bullying victimization at grade six versus anxiety and depression at grade eight. We located the study in our electronic searches because at the follow-up, children also filled the Bully Survey (Swearer & Paulk, 1998). The children did not fill in a bullying questionnaire in grade six. Subsequently, given the research questions of our review, the study is excluded because of the cross-sectional character of the data of interest.

⁶ To be more precise, in a 6-month follow-up, the authors (Shelley, 2009; Shelley & Craig, 2010) show the association of bullying (perpetration and victimization) with various outcomes, including a depressive attribution style (unadjusted effect sizes). The authors also show results from step-wise regressions, with bullying and depression at Time 1 being regressed on victimization at Time 2 (all scales distributed at both times). Subsequently, we report only an unadjusted effect size for these data while the adjusted effect sizes are excluded.

⁷ Moon *et al.* (2011) present data for a one-year follow-up study in Korea. The study findings are not relevant to our review, though, since the measures of interest (i.e. depression and anger) are taken at Time 1 and their relationship with bullying at Time 2 is examined (see Moon *et al.*, 2011: 16 - 18; see tables 3 and 4).

lished papers also followed our analytical approach (e.g. Fekkes et al., 2006, with depression as an outcome measure). Various evaluators of anti-bullying programmes provided relevant data (e.g. Dorothy Espelage for the Multimedia Violence Prevention Study⁸; Caroline Hunt for the Confident Kids Programme; Christina Salmivalli for the KiVa Programme; and Rolf Sandell for the SET Project), while others could not carry out the requested data analyses (e.g. the S.S.GRIN Programme⁹, the Beyond Bullying Secondary Programme¹⁰ and the Owning up Programme¹¹).

6. Study participants are school-aged children in the community and exposure to bullying (perpetration and victimization) had to specify school years. We have excluded, therefore, the paper on depression by Jordanova et al. (2007)¹² after confirmation by Robert Stewart that exposure to bullying under ‘lifetime events’ did not necessarily concentrate on school bullying victimization. We did, however, include retrospective studies, in which the study participants are adults and in which a retrospective measure of exposure to school bullying is related to outcome measures of interest (i.e. concurrent depression or offending).
7. The report has quantitative data that allow calculation of an effect size. For example, the study on depression by Carlisle and Rofes (2007) was based on qualitative data and was excluded.
8. We included published and unpublished reports of the literature including books (e.g. for offending: Haas, 2001; Olweus, 1993a) and book chapters (e.g. for depression: Olweus, 1993c, 1994b; for offending: Olweus, 1993b), journal articles, Masters or PhD theses (e.g. for depression: Blais, 2008; Grills, 2003; Parada, 2006; Singer, 2002; Taylor, 2006; and for offending: Wong, 2009) and conference presentations (e.g. Lösel et al., 2008). Data were also obtained via email communications with Principal Investigators of major longitudinal studies (see later).

⁸ The study shows specific results on school bullying and not just aggression. Dorothy Espelage provided the zero-order correlation coefficient for Time 1 bullying perpetration versus Time 2 depression for the sixth graders who were part of the control group (email: December 3, 2010). Adjusted effect sizes could not be provided.

⁹ Email communication with Melissa DeRosier, January 4, 2011.

¹⁰ We were unable to find the address for correspondence of Roberto Parada. Various emails have been sent to Herbert Marsh since January 15, 2011, but we did not receive any response.

¹¹ We were not able to find the email address of Randie Taylor at all.

¹² Email communication with Dr Robert Stewart: January 13, 2011.

Some criteria for exclusion of reports were as follows:

1. Bullying perpetration/victimization is a sub-scale of a peer victimization/ aggression scale and effect sizes are not shown for the bullying subscale.
2. The outcome measure (i.e. depression and offending) is part (i.e. a subscale) of a wider theoretical construct (e.g. ‘overall health’ or ‘antisocial behaviour’) and effect sizes are not shown for each subscale (e.g. for depression: Farrington & Ttofi, 2011)¹³. This is the main reason why specific reports relating to the E-Risk study (i.e. Arseneault, 2011, Shakoor et al., 2011) were excluded: both depression and delinquency were part of wider theoretical constructs (i.e. internalizing and externalizing problem behaviour).
3. The outcome of interest (i.e. depression or offending) is used as a moderator between school bullying and another outcome (e.g. for depression: Hidaka & Operario, 2006¹⁴; Roeger et al., 2010¹⁵) or simply as another independent predictor alongside bullying (e.g. Nrugham et al., 2008)¹⁶.
4. Study participants attend institutions for incarcerated or institutionalized youth. Three independent studies on the link between bullying perpetration and offending in the Netherlands (Bijeveld et al., 2011) were excluded because of this feature. Similarly, a Finnish study on bullying perpetration by Luukkonen et al. (2011)¹⁷, with offending as the outcome measure was excluded since study participants were inpatient adolescents.
5. Finally, studies that consistently used the term bullying while it was clear (from the description of the variables) that they were actually concerned with general aggression/victimization were excluded (e.g. Azzuzi & Killias, 2010).¹⁸

¹³ Depression is part of the composite measure of life success.

¹⁴ Hidaka and Operario (2006) show unadjusted effect sizes for bullying victimization at school (retrospective measure) versus attempted suicide among GBQ Japanese men based on bivariate logistic regressions. They also give adjusted effect sizes based on multivariate logistic regressions after controlling for various measures, including depression (see table 2: 965).

¹⁵ For this longitudinal retrospective study, Roeger et al. (2010) show adjusted effect sizes for bullying versus suicidal ideation after controlling for depression (see table 3: 732).

¹⁶ Depression and bullying as independent predictors of suicidal acts (Nrugham et al., 2008: 37 – 38; see tables 2 and 3; see also Nrugham, 2010).

¹⁷ Email communication with Anu-Helmi Halt, August 22, 2011.

¹⁸ The retrospective longitudinal study (unpublished manuscript) included a clearly stated measure of ‘being bullied’ (single item) but not a clear measure of bullying perpetration (named as active bullying but actually dealing with aggressive behaviour). In a later published version of the paper (Staubli & Killias, 2011) results are presented only for bullying victimization.

2.2 Searching Strategies

(a) We started by searching for the names of established researchers in the area of bullying prevention research (e.g. Australia, K. Rigby; England, P.K. Smith; Finland, C. Salmivalli; Italy, E. Menesini; Norway, D. Olweus; Spain, R. Ortega), since many prevention programmes are multi-component and target wider behavioural problems. This searching strategy was used in different databases in order to initially obtain as many relevant studies in different journals as possible.

(b) We then searched using several keywords in different databases. In total, we carried out the same searching strategies in 19 electronic databases (see Table 2). In all databases, the same key words (covering more outcomes than those analysed in the present report) were used with different combinations:

- School bully; school bullies; school bully-victims; school bullying victimization

AND

- psychosomatic; health outcomes; suicidal ideation; eating disorders; psychiatric symptoms; neuroticism; psychosocial; physical health; mental health; self-harm; delinquency; criminality; psychosis; psychometric; trauma; disorders; clinicians; interns; pain; illness; self-injurious; stress; clinical; distress; offending; vandalism; theft; arson/ fire-setting; depression; anxiety; violence; aggression.

Table 2. List of Databases Searched

Australian Criminology Database (CINCH)
Australian Education Index
British Education Index
Cochrane Controlled Trials Register
C2-SPECTR
Criminal Justice Abstracts
Database of Abstracts of Reviews of Effectiveness (DARE)
Dissertation Abstracts
Educational Resources Information Clearinghouse (ERIC)
Ethos-Beta
EMBASE
Google Scholar
Index to Theses Database
MEDLINE
National Criminal Justice Reference Service (NCJRS)
PsychInfo/Psychlit
Sociological Abstracts
Social Sciences Citation Index (SSCI)
Web of Knowledge

- (c) In addition, 63 journals have been hand searched. Table 3 gives a list of the journals that we have hand-searched, either online or in print. Furthermore, beginning in 2009, we subscribed to the Zetoc database, which covers tables of contents of journals from 1993 to date and is updated on a daily basis. This email alerting service enabled us to keep up-to-date with relevant new articles in many journals that the University of Cambridge does not subscribe to. For the Zetoc email alerting service, the general key word of ‘bullying’ was used for either the title of the abstract. In this way, we were able to obtain and screen all relevant papers dealing with bullying and obtain those relevant through interlibrary loans.
- (d) A stipulation was made that the title or abstract of each paper would have to include one of the essential key words that were searched. However, some book chapters, mainly from edited volumes, were included even though their titles and/or abstracts (if provided) did not include any of our key words.
- (e) We have contacted the principal investigators of a large number of prospective longitudinal studies across the world and asked them to carry out new data analyses on the topic of school bullying and its outcomes. We have explained the aims of our review and our analytical strategy for the meta-analysis. Authors were sent the Murray et al. (2009) paper on drawing conclusions about causes from systematic reviews of risk factors and were given guidelines about how to provide unadjusted and adjusted effect sizes. In total, scholars and research teams from 29 longitudinal studies have provided unpublished data (see table 4). Specifically, we have received results from data analyses carried out by 24 researchers. We have also received raw datasets for the following three studies: a) Coimbra Prospective Longitudinal Study/Young Cohort; b) Coimbra Prospective Longitudinal Study/Intermediate Cohort¹⁹; and c) Seattle Social Development Study²⁰. Two more researchers²¹ have provided initial data analyses and raw datasets for completion of analyses by our research team. We are hopeful that we might be able to complete analyses of the above five studies in due course. Of the 24 studies with complete data analyses, 15 are presented in two special issues that have been organised by David Farrington, Friedrich

¹⁹ For both Portuguese studies, data received via email communication with Antonio Fonseca, May 20, 2010.

²⁰ Email communication with Karl Hill, November 4, 2010.

²¹ For the Oregon Youth Study, Deborah Capaldi (email dated March 23, 2010) provided partial correlation coefficients after controlling for the Antisocial Behaviour Construct Score. Bullying and aggression may be confounded with a general antisocial behaviour construct, so we have excluded the study. For the New York Longitudinal Study (email communication of David Farrington with Patricia Cohen, February 16, 2010), Patricia Cohen provided initial data analyses along with raw data so that further analyses could be carried out.

Table 3. List of Journals Searched

Archives of Pediatrics and Adolescent Medicine	International Journal of Behavioral Medicine
Aggression and Violent Behavior	International Journal on Violence and Schools
Aggressive Behavior	Intervention in School and Clinic
American Journal of Psychiatry	Japanese Journal of Educational Psychology
Australian Journal of Education	Journal of Adolescent Health
Australian Journal of Educational and Developmental Psychology	Journal of the American Medical Association
British Journal of Clinical Psychology	Journal of Behavioral Medicine
British Journal of Developmental Psychology	Journal of Child Psychology and Psychiatry
British Journal of Educational Psychology	Journal of Educational Psychology
British Journal of Psychiatry	Journal of Emotional Abuse
British Medical Journal	Journal of Experimental Criminology
Canadian Journal of School Psychology	Journal of Interpersonal Violence
Child Abuse and Neglect	Journal of Pediatric Psychology
Child Development	Journal of Psychosomatic Research
Child Psychiatry and Human Development	Journal of School Violence
Clinical Psychology Review	Journal of School Health
Criminal Justice and Behavior	Journal of Youth and Adolescence
Crisis-The journal of Crisis Intervention and Suicide Prevention	Justice Quarterly
Developmental Psychology	Pastoral Care in Education
Development and Psychopathology	Psychological Medicine
Deviant Behavior	Psychology, Crime and Law
Educational Psychology	Psychology Health and Medicine
Educational Psychology in Practice	Psychology in the Schools
Educational Psychology Review	Scandinavian Journal of Psychology
Educational Research	School Psychology International
European Journal of Public Health	School Psychology Review
Health Education Journal	Studies in Educational Evaluation
Health Promotion International	Swiss Journal of Psychology
Health Education Research	Trauma, Violence and Abuse
Injury Prevention	Victims and Offenders
International Journal of Behavioral Development	Violence and Victims Youth and Society

Lösel and Maria Ttofi in two peer-reviewed journals, namely: in the *Journal of Aggression, Conflict and Peace Research* (Ttofi et al., 2011a) and in *Criminal Behaviour and Mental Health* (Farrington et al., 2011b). In both special issues, the support of the Swedish National Council for Crime Prevention is highlighted.

Table 4. List of Longitudinal and Intervention/Follow-up Studies for which New Data Analyses were Provided by Researchers

-
1. Australian Temperament Project (Renda et al., 2011)²²
 2. Cambridge Study in Delinquent Development (Farrington, 1993; Farrington & Ttofi, 2011)
 3. Christchurch Health and Development Study (Gibb et al., 2011)
 4. Confident Kids Program (Berry & Hunt, 2009)²³
 5. Dunedin Longitudinal Study (Moffitt et al., 2010)²⁴
 6. Edinburgh Study of Youth Transitions and Crime (Barker et al., 2008; McVie, 2010; Smith & Ecob, 2007)
 7. E-Risk Longitudinal Study (Shakoor et al., 2011; Arsenaault, 2011)²⁵
 8. Erlangen-Nuremberg Development and Prevention Study (Lösel and Bender, 2010)²⁶
 9. Erlangen-Nuremberg Longitudinal Study of Bullying (Bender & Lösel, 2011; Lösel & Bliesener, 2003; Lösel et al., 2008)
 10. International Youth Development Study (Hemphill et al., 2011; Patton et al., 2008)
 11. Japanese Longitudinal Study (Nishino et al., 2009; Nishino, 2010²⁷ /email; Nishino et al., 2011)
 12. KiVa Anti-Bullying Programme (Salmivalli, 2010)²⁸
 13. Mater-University of Queensland Study of Pregnancy and its Outcomes (McGee et al., 2011)
 14. Metropolitan Area Child Study (Henry et al., 2010)²⁹
 15. Metropolitan Area Child Study (Henry et al., 2010)
 16. Multimedia Violence Prevention Study (Espelage et al., 2001)³⁰

²² For depression, results were provided via email communication with Jennifer Renda (July 16, 2010).

²³ The authors have provided standardized regression coefficients for bullying victimization at baseline (before the implementation of the programme) versus depression at the follow-up for the control group only (email communication with Caroline Hunt, May 26, 2010). Bullying victimization was a continuous variable.

²⁴ Email communication with Retate Houts, July 22, 2010.

²⁵ Adjusted effect sizes provided by Louise Arsenaault via email communication (January, 21, 2011).

²⁶ Email communication with Friedrich Lösel, December 31, 2010.

²⁷ Email communication with Yasuyo Nishino, March 30, 2010.

²⁸ Results given via email communication with Christina Salmivalli (March 29, 2010).

²⁹ Results obtained via email communication with David Henry (July 16, 2010). The two reports are based on two independent cohorts.

³⁰ The study shows specific results on school bullying and not just aggression. Dorothy Espelage has provided the zero-order correlation coefficient for Time 1 bullying perpetration versus Time 2 depression for the sixth graders who were part of the control group (email: December 3, 2010).

17. Pittsburgh Youth Study (Farrington et al., 2011a; White & Loeber, 2008)
18. Raising Healthy Children Project (Kim et al., 2011)
19. SET Project (Kimber et al., 2008a, 2008b)³¹
20. SNAP Under 12 Outreach Project (Jiang et al., 2011)
21. Seven Schools Longitudinal Study (Kendrick & Stattin, 2010³²; Ozdemir & Stattin, 2011)
22. Swedish Community Samples (Olweus 1991; 1993a, b, c; 1994a, b; 1997; 2011)
23. Swiss Federal Survey of Army Recruits of 1997 (Azzuzi & Killias, 2010; Haas, 2001; Staubli & Killias, 2011)
24. z-proso Longitudinal Study (Averdijk et al., 2011)

Note: All relevant papers for each study are presented and not just the most recent ones. Datasets for further analyses have been provided for: a) the Coimbra Prospective Longitudinal Study/Young Cohort; b) the Coimbra Prospective Longitudinal Study/Intermediate Cohort; c) the Seattle Social Development Study; d) the Oregon Youth Study and; e) the New York Longitudinal Study (see relevant text in report).

In the special issue of *JACPR*, a meta-analysis is presented on the association of bullying victimization with later depression (Ttofi et al., 2011b). In the special issue of *CBMH*, a meta-analysis is presented on the association of bullying perpetration with later offending (Ttofi et al., 2011c). In the current report, we go beyond the work of these special issues by presenting further results from two new meta-analyses: (a) for bullying perpetration versus later depression; and (b) for bullying victimization versus later offending. Furthermore, moderator analyses that may explain variations in adjusted effect sizes in the two new meta-analytic reviews will be presented. Finally, we have included new longitudinal studies in the present report (e.g.: Sourander et al., 2011 study on criminality; Vaillancourt et al., 2011, available online in August 2011, four months after our two special issues were available online).

2.3 Screening of Reports

A total number of 661 reports that were concerned with the association of bullying (perpetration and/or victimization) with internalizing and/or externalizing problems were found. Table 5 presents the number of reports with relevant data based on both cross-sectional and longitudinal (prospective or retrospective) studies. Reports were screened based on a relevance scale that we have constructed, reducing the final number of included reports (i.e. category 3) to 462 (69.9%). Category 3 reports presented data on the association of bullying (perpetration or victimization) with various outcome measures (not just with depression or offending). Of the 462 re-

³¹ Special data analyses results (only adjusted effect sizes) provided via email communication with Rolf Sandell (email: March 19, 2010).

³² Email communication with Kristin Kendrick; February 22 and 26, 2010.

ports, 337 presented data analyses based on cross-sectional data (72.9%), 96 were based on prospective longitudinal data (20.8%), while 29 included longitudinal data analyses based on a retrospective measure of school bullying (retrospective longitudinal studies; 6.3%).

Table 5. Categorization of 661 Reports Based on their Relevance to the Systematic Review

Category 1. Excluded reports: Reports with qualitative data or theoretical papers such as narrative reviews (N = 125; 18.9%)

Category 2. Excluded reports: Reports with relevant data in which either the predictors or the outcome measures consist of a sub-scale of a wider instrument, and with no statistical data presented for the subscales of interest (N = 22; 3.3%)

Category 3. Included reports: Reports with data on the association of bullying perpetration or victimization with internalizing or externalizing problems (N = 462; 69.9%)

Category 4. Includable reports: Reports with relevant data which are potentially includable if further information could be received (e.g. statistical measures for obtaining an effect size are missing; or reports needing translation) (N = 47; 7.2%)

Category 5. Relevant raw data from prospective longitudinal studies needing further analyses (N = 5; 0.8 %)

An effort was made to include all types of reports, including book chapters (N = 24, 3.6% of all reports; 4.3% of category 3), unpublished Ph.D. or Masters theses (N = 41, 6.2% of all reports; 6.5% of category 3), and technical or other reports (e.g. conference papers or data obtained via email communication: N = 27, 4.1% of all reports; 4% of category 3). The majority of reports were presented in peer-reviewed journals (N = 569, 86.1% of all reports; 85.3% of category 3).

The total number of reports that addressed the link between school bullying (perpetration or victimization) and later depression has increased markedly over time, as shown in figure 1. An increasing trend is also shown for studies of the association between school bullying (perpetration and victimization) and later offending, increasing especially for analyses of longitudinal studies in the most recent time period (figure 2)³³. The total time period was divided into 5-year chunks apart from the period covering 1971 to 1992, since only 10 reports were located during this first period³⁴.

³³ Plots of time trends are not limited to category 3 studies.

³⁴ Specifically: 1 report in 1971, 2 in 1987, 1 in 1989, 3 in 1990, 1 in 1991 and 2 in 1992.

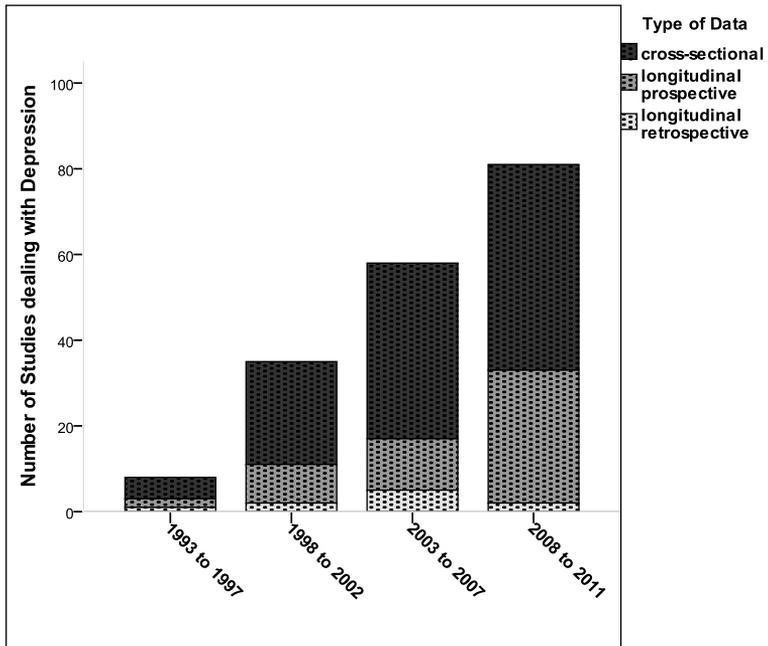


Figure 1. Number of Reports on Bullying and Depression within Year Periods

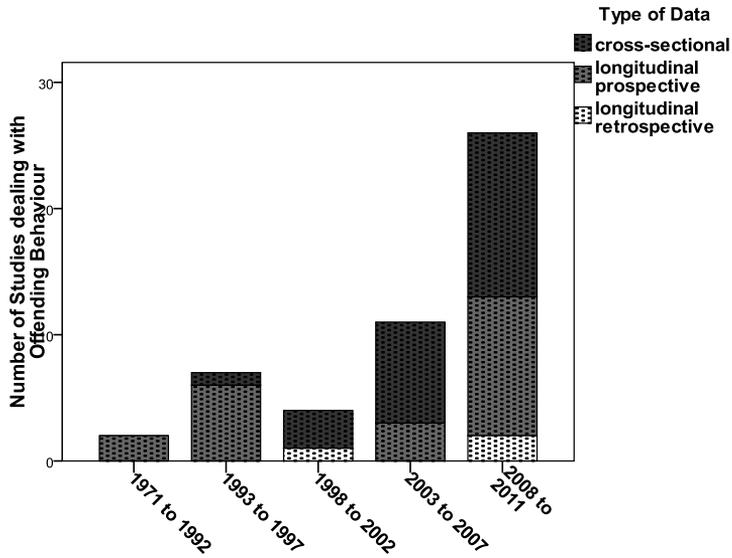


Figure 2. Number of Reports on Bullying and Offending within Year Periods

2.4 Included and Excluded Studies

A total number of 48 reports from 29 longitudinal studies presented data on the long-term association of school bullying (perpetration and/or victimization) with offending in adolescence or young adulthood, as shown in table 6. A total number of 75 reports from 49 longitudinal studies presented data on the long-term association of school bullying with depression in adolescence or young adulthood, as shown in Table 7. Both tables list included and excluded studies, the number of published or unpublished reports related to each study, and the type of longitudinal data (prospective, retrospective or follow-up interventions) in each study.

Table 6. 48 Reports on Offending from 29 Longitudinal Studies

(A) Included Studies:

Australian Temperament Project (Renda et al., 2011); longitudinal prospective

=> police/court contact based on self-reports at age 21.5; bullying at age 13.5; controlling for 7 covariates

=> combined property damage and shoplifting (separate items) based on self-reports at age 23.5; victimization at age 13.5; controlling for 20 covariates

Cambridge Study in Delinquent Development (Farrington, 1993; Farrington & Tofi, 2011); longitudinal prospective

=> offending based on convictions (official record data) at age 17.5; bullying at age 14; controlling for 20 covariates

Christchurch Health and Development Study (Gibb et al., 2011); longitudinal prospective

=> combined property offending and arrest/conviction (separate measures) based on self-reports at age 23; bullying at age 11.75; controlling for 16 covariates

=> combined property offending and arrest/conviction (separate measures) based on self-reports at age 23; victimization at age 14; controlling for 14 covariates

Edinburgh Study of Youth Transitions and Crime (Barker et al., 2008; McVie, 2010; Smith & Ecob, 2007); longitudinal prospective

=> combined property theft and damage (separate items) based on self-reports at age 14; bullying at age 13; controlling for 10 covariates

=> combined property theft and damage (separate items) based on self-reports at age 14; victimization at age 13; controlling for 10 covariates

Erlangen-Nuremberg Development and Prevention Study (Lösel and Bender, 2011³⁵); longitudinal prospective and intervention study

=> combined self-reported and mother-reported delinquency (separate measures) for offending at age 13.7; bullying at age 9; controlling for 5 covariates

=> combined self-reported and mother-reported delinquency (separate measures) for offending at age 13.7; victimization at age 9; controlling for 5 covariates

Erlangen-Nuremberg Longitudinal Study of Bullying (Bender & Lösel, 2011; Lösel & Bliesener, 2003; Lösel et al., 2008); longitudinal prospective

=> delinquency based on self-reports (total GDFB scale) at age 24.64; bullying at age 15.54; controlling for 3 covariates

=> delinquency based on self-reports (total GDFB scale) at age 24.64; victimization at age 15.54; controlling for 3 covariates

³⁵ Email communication with Friedrich Lösel, December 31, 2010.

From a Boy to a Man Finnish Longitudinal Study (Sourander et al., 2006, 2007a) and the Nationwide Finnish 1981 Birth Cohort Study (Sourander et al., 2011); longitudinal prospective

=> criminal offences based on official records at age 24.5; bullying at age 8; controlling for 2 covariates

=> criminal offences based on official records at age 24.5; victimization at age 8; controlling for 2 covariates

International Youth Development Study (Hemphill et al., 2011); longitudinal prospective

=> theft based on self-reports at age 16.9; bullying at age 14.4; controlling for 8 covariates

=> theft based on self-reports at age 16.9; victimization at age 14.4; controlling for 8 covariates

Japanese Longitudinal Study (Nishino et al., 2009; Nishino, 2010³⁶/email; Nishino et al., 2011); longitudinal prospective/short-term follow-up study

=> combined shoplifting and vehicle theft (separate measures) based on self-reports at age 12.92; bullying measured at age 12.5; controlling for 4 covariates

=> combined shoplifting and vehicle theft (separate measures) based on self-reports at age 12.92; victimization measured at age 12.5; controlling for 4 covariates

Jyvaskyla Longitudinal Study in Finland (Pulkkinen & Tremblay, 1992); longitudinal prospective

=> total criminal records ('all registers') for offending based on official records; unadjusted effect sizes only

Mater-University of Queensland Study of Pregnancy and its Outcomes (McGee et al., 2011); longitudinal prospective

=> delinquency (single item) at age 21 based on self-reports; victimization at age 14; controlling for 2 covariates

Metropolitan Area Child Study (Henry et al., 2010); Study 1; longitudinal prospective

=> delinquency based on self-reports at age 10; bullying at age 8; controlling for 4 covariates

=> delinquency based on self-reports at age 10; victimization at age 8; controlling for 4 covariates

Metropolitan Area Child Study (Henry et al., 2010); Study 2; longitudinal prospective

=> delinquency based on self-reports at age 13; bullying at age 11; controlling for 4 covariates

=> delinquency based on self-reports at age 13; victimization at age 11; controlling for 4 covariates

Montreal Longitudinal Study (Haapasalo et al., 2000; Tremblay & Haapasalo, 1998; Pulkkinen & Tremblay, 1992); longitudinal prospective

=> delinquency based on self-reports at age 11; bullying at age 6.23; controlling for 1 covariate

National Longitudinal Survey of Youth 1997 (Wong, 2009); longitudinal prospective

=> combined arrest, theft, vandalism and other property crime (4 separate items) based on self-reports at age 14.34; victimization at age 12; controlling for 20 covariates

Pittsburgh Youth Study (Farrington et al., 2011a); longitudinal prospective

=> Delinquency based on self-reports at age 14.27; bullying at age 10.98; controlling for 10 covariates

Raising Healthy Children Project (Kim et al., 2011); longitudinal prospective and intervention study

=> violent offending based on self-reports at age 21.52; bullying at age 11.5; controlling for 6 covariates

³⁶ Email communication with Yasuyo Nishino, March 30, 2010.

Seven Schools Longitudinal Study (Kendrick & Stattin, 2010; email³⁷); longitudinal prospective

=> property crimes based on self-reports; unadjusted effect sizes only

SNAP Under 12 Outreach Project (Jiang et al., 2011); longitudinal prospective and intervention study

=> offending based on official records at age 17.99; bullying at age 9.5; controlling for 5 covariates

Swedish Community Samples (Olweus 1991; 1993a, b, c; 1994a, b; 1997; 2011); longitudinal prospective

=> offending based on official records; unadjusted effect sizes only

Swiss Federal Survey of Army Recruits of 1997 (Azzuzi & Killias, 2010; Haas, 2001; Staubli & Killias, 2011); longitudinal retrospective

=> offending based on four separate self-reported items (knifed, strangled, shot with gun, shot with firearm) at age 19.5; victimization at age 8.5; unadjusted effect sizes only

(B) Excluded Studies or Specific Reports of Included Studies:

E-Risk Longitudinal Study (Arseneault, 2011; Bowes et al., 2009, 2010; Shakoor et al., 2011);³⁸

Five-month follow-up of English Students (Boulton et al., 2010)

STUDY-70 Project: Follow-up Study of Finnish Inpatient Adolescents (Luukkonen et al., 2011)

Official Records Follow-Up Study in the Netherlands; Study 1 (Bijleveld et al., 2011)

Official Records Follow-Up Study in the Netherlands; Study 2 (Bijleveld et al., 2011)

Official Records Follow-Up Study in the Netherlands; Study 3 (Bijleveld et al., 2011)

Pittsburgh Youth Study (White & Loeber, 2008)

Project GANGFACT (Holmes et al., 1998)

Two-year Follow-up Study of London Children (Smith et al., 2004)

Table 7. 75 Reports on Depression from 49 Longitudinal Studies

(A) Included Studies:

Adolescent Mental Health Cohort Study (Kaltiala-Heino et al., 2010); longitudinal prospective

=> depression at age 17; bullying perpetration at age 15; controlling for 4 covariates

=> depression at age 17; bullying victimization at age 15; controlling for 4 covariates

Australian Temperament Project (Renda et al., 2011)³⁹; longitudinal prospective

=> depression at age 23.5; bullying perpetration at age 13.5; controlling for 20 covariates

=> depression at age 23.5; bullying victimization at age 13.5; controlling for 20 covariates

Christchurch Health and Development Study (Gibb et al., 2011); longitudinal prospective

=> depression at age 23; bullying perpetration at age 11.75; controlling for 16 covariates

=> depression at age 23; bullying victimization at age 14; controlling for 14 covariates

³⁷ Email communication with Kristin Kendrick; February 22 and 26, 2010.

³⁸ Adjusted effect sizes provided by Louise Arseneault via email communication (January, 21, 2011).

³⁹ For depression, results were provided via email communication with Jennifer Renda (July 16, 2010).

Confident Kids Program (Berry & Hunt, 2009); follow-up/intervention study⁴⁰

=> depression at age 13.21; bullying victimization at age 13.04; unadjusted effect size only

Danish Longitudinal Health Behaviour Study, Young Cohort (Due et al., 2009)⁴¹; longitudinal retrospective

=> depression at age 27; bullying victimization at age 15; unadjusted effect sizes only

Danish Longitudinal Retrospective Study (Lund et al., 2008); longitudinal retrospective

=> depression at age 41; bullying victimization at age 18; controlling for 2 covariates

Dunedin Longitudinal Study (Moffitt et al., 2010)⁴²; longitudinal prospective

=> depression at age 32; bullying perpetration at age 8; controlling for 5 covariates

Dutch Anti-Bullying Programme (Fekkes et al., 2006); follow-up/intervention study

=> depression at age 10.5; bullying victimization at age 10; unadjusted effect sizes only

Edinburgh Study of Youth Transitions and Crime (McVie, 2010); longitudinal prospective

=> depression at age 14; bullying perpetration at age 13; controlling for 10 covariates

=> depression at age 14; bullying victimization at age 13; controlling for 10 covariates

Erlangen-Nuremberg Development and Prevention Study (Lösel & Bender, 2011); longitudinal prospective and intervention study

=> depression at age 13.7; bullying perpetration at age 9; controlling for 5 covariates

=> depression at age 13.7; bullying victimization at age 9; controlling for 5 covariates

Erlangen-Nuremberg Longitudinal Study of School Bullying (Bender & Lösel, 2011; Lösel & Bliesener, 2003; Lösel et al., 2008); longitudinal prospective

=> depression at age 24.64; bullying perpetration at age 15.54; controlling for 3 covariates

=> depression at age 24.64; bullying victimization at age 15.54; controlling for 3 covariates

European TMR Network Project (Singer, 2002); longitudinal retrospective

=> depression at age 23.5; bullying victimization at age 11; controlling for 7 covariates; adjusted effect sizes only

Follow-Up Study in Canada (Vaillancourt et al., 2011); longitudinal prospective

=> depression at age 13.25; bullying victimization at age 12.25; unadjusted effect sizes only

'From a Boy to a Man' Finnish Longitudinal Study (Haavisto et al. 2004; Klomek et al. 2008; Sourander et al. 2007b) ; part of the Nationwide 1981 Finnish Longitudinal Study (Sourander et al. 2009) ; longitudinal prospective

=> depression at age 18; bullying perpetration at age 8; controlling for 1 covariate

=> depression at age 18; bullying victimization at age 8; controlling for 1 covariate

Gatehouse Project (Bond et al., 2001); follow-up/intervention study⁴³

=> depression at age 14; bullying victimization at age 13; controlling for 5 covariates

Health 2000 Project (Pirkola et al., 2005); longitudinal retrospective⁴⁴

=> depression reports at age 47; bullying victimization at age 11; controlling for 11 covariates

⁴⁰ The authors have provided standardized regression coefficients for bullying victimization at baseline (before the implementation of the programme) versus depression at the follow-up for the control group only (email communication with Caroline Hunt, May 26, 2010). Bullying victimization was a continuous variable.

⁴¹ Further to our email correspondence (January 14, 2011), Pernille Due has agreed to provide adjusted effect sizes in due course. For the moment, we can only report the results presented in the published paper.

⁴² Email communication with Retate Houts, July 22, 2010.

⁴³ The authors have controlled for the implementation group.

⁴⁴ In a previous publication (Ttofi et al., 2011b), we have mistakenly indicated three confounds controlled for in the adjusted effect size instead of eleven.

International Youth Development Study (Hemphill et al., 2011; Patton et al., 2008); longitudinal prospective

=> depression at age 16.9; bullying perpetration at age 14.4; controlling for 8 covariates
=> depression at age 16.9; bullying victimization at age 14.4; controlling for 8 covariates

Japanese Longitudinal Study (Nishino et al., 2009; Nishino et al., 2011); longitudinal prospective

=> depression at age 12.92; bullying perpetration at age 12.5; controlling for 4 covariates
=> depression at age 13.81; bullying victimization at age 12.5; controlling for 4 covariates

KiVa Anti-Bullying Programme (Salmivalli, 2010)⁴⁵; follow-up/intervention study

=> depression at age 10.5; bullying perpetration at age 9.5; unadjusted effect sizes only
=> depression at age 10.5; bullying victimization at age 9.5; unadjusted effect sizes only

Longitudinal Retrospective Study at the Mood Disorders Unit Outpatient Depression Clinic in Sydney, Australia (Gladstone et al., 2006); longitudinal retrospective

=> depression at age 43; bullying victimization at age 10.5; unadjusted effect sizes only

Longitudinal Retrospective Study of Adult Twin Pairs (Gladstone & Parker, 2006); longitudinal retrospective

=> depression at age 40.7; bullying victimization at age 8.5; unadjusted effect sizes only

Longitudinal Retrospective Study of American University Students (Roth et al., 2002); longitudinal retrospective

=> depression at age 19.36; bullying victimization at age 12; controlling for 1 covariate

Longitudinal Retrospective Study of English GBQ men (Rivers, 1999, 2001; Rivers & Cowie, 2006); longitudinal retrospective

=> depression at age 28; bullying victimization at school (not specified); unadjusted effect sizes only

Longitudinal Retrospective Study of Japanese University Students (Matsui et al., 1996)⁴⁶; longitudinal retrospective

=> depression at age 19.4; bullying victimization at age 13.5; unadjusted effect sizes only

Mater-University of Queensland Study of Pregnancy and its Outcomes (McGee et al., 2011); longitudinal prospective

=> depression at age 20.9; bullying victimization at age 13.9; controlling for 3 covariates

Metropolitan Area Child Study (Henry et al., 2010)⁴⁷; Study 1; longitudinal prospective

=> depression at age 10; bullying perpetration at age 8; controlling for 4 covariates
=> depression at age 10; bullying victimization at age 8; controlling for 4 covariates

Metropolitan Area Child Study (Henry et al., 2010); Study 2; longitudinal prospective

=> depression at age 13; bullying perpetration at age 11; controlling for 4 covariates
=> depression at age 13; bullying victimization at age 11; controlling for 4 covariates

Multimedia Violence Prevention Study (Espelage et al., 2001)⁴⁸; follow-up/intervention study

=> depression at age 13.07; bullying perpetration at age 12.74; unadjusted effect sizes only

⁴⁵ Results given via email communication with Christina Salmivalli (March 29, 2010).

⁴⁶ In the JACPR meta-analysis paper, we mistakenly indicate that the report includes adjusted effect sizes only. We did not include the adjusted effect sizes in the JACPR meta-analysis; please see relevant text.

⁴⁷ Results obtained via email communication with David Henry (July 16, 2010). The two reports are based on two independent cohorts.

⁴⁸ The study shows specific results on school bullying and not just aggression. Dorothy Espelage has provided the zero-order correlation coefficient for Time 1 bullying perpetration versus Time 2 depression (email: December 3, 2010).

Pittsburgh Youth Study (Farrington et al., 2011a); longitudinal prospective

=> depression at age 14.27; bullying victimization at age 10.98; controlling for 10 covariates

SET Project (Kimber et al., 2008a, 2008b)⁴⁹; follow-up/intervention study

=> depression at age 14.5; bullying victimization at age 13.5; controlling for 3 covariates; adjusted effect sizes only

Seven Schools Longitudinal Study (Ozdemir & Stattin, 2011); longitudinal prospective

=> depression at age 14.49; bullying perpetration at age 13.2; controlling for 2 covariates

=> depression at age 14.49; bullying victimization at age 13.2; controlling for 2 covariates

Six-month follow-up study in Canada (Shelley, 2009; Shelley & Craig, 2010)⁵⁰; longitudinal prospective

=> depression at age 11.5; bullying perpetration at age 11; unadjusted effect size only

=> depression at age 11.5; bullying victimization at age 11; unadjusted effect size only

Swedish Community Samples (Olweus, 1993c, 1994b); longitudinal prospective

=> depression at age 23; bullying victimization at age 16; unadjusted effect sizes only

z-proso Longitudinal Study (Averdijk et al., 2011); longitudinal prospective and intervention study

=> depression at age 11; bullying victimization at age 8; controlling for 11 covariates

(B) Excluded Studies or Specific Reports of Included Studies:

Beyond Bullying Secondary Programme (Marsh et al., 2004; Parada, 2006; Parada et al., 2008)

Birth to Ten Longitudinal Study in South Africa (Barbarin, 1999)

British National Survey of Psychiatric Morbidity (Jordanova et al., 2007)⁵¹

Cambridge Study in Delinquent Development (Farrington & Ttofi, 2011)

E-Risk Longitudinal Study (Arsenault, 2011; Bowes et al., 2009, 2010; Shakoor et al., 2011)⁵²

Finnish Cohort Longitudinal Study (Kumpulainen & Rasanen 2000; Kumpulainen & Roine, 2002; Kumpulainen et al., 2000; Kumpulainen et al. 2001; Klomek et al. 2009; Sourander et al., 2000)

Health Omnibus Survey in South Australia (Roeger et al., 2010)

Longitudinal Retrospective Study of English GBQ men (Rivers, 2004)

Longitudinal Retrospective Study of Japanese GBQ men (Hidaka & Operario, 2006)

Longitudinal Study in Korea (Moon et al., 2011)

Norwegian Follow-Up Study (Nrugham, 2010; Nrugham et al., 2008)

Owning Up Bullying Prevention Programme (Taylor, 2006)

Pilot Study of adult males in US, UK and Australia (Carlisle & Rofes, 2007)

S.S.GRIN Intervention Study (DeRosier, 2004, 2007; DeRosier & Marcus, 2005)

Three-year Follow-up Study in Australia (Rigby, 1999, 2001)⁵³

Two-year Follow-up of Virginia Students (Grills, 2003)

Video Game Violence Follow-Up Study (Ferguson, 2011)⁵⁴

⁴⁹ Special data analyses results (only adjusted effect sizes) provided via email communication with Rolf Sandell (email: March 19, 2010).

⁵⁰ Only unadjusted effect sizes are included. The adjusted effect sizes for the study were excluded for reasons explained above (see relevant explanation in part 2.3 on inclusion/exclusion criteria).

⁵¹ Excluded further to email correspondence with Robert Stewart (January, 13, 2011).

⁵² Adjusted effect sizes provided by Louise Arsenault via email communication (January, 21, 2011).

⁵³ Bullying victimization, anxiety and depression were part of the baseline data. Questions on depression, but not on anxiety, were excluded from the follow-up period upon request from the schools.

⁵⁴ Depression at Time 1 and 2 are used as predictors for bullying at Time 2 (see: Ferguson, 2011: 386, table 2).

Table 8. Summary Table of Included Reports

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Adolescent Mental Health Cohort Study</i> (Kaltiala-Heino et al. 2010)	Finland	3278 students	No: 4 * Depression at T1 * Child's age * Parental education * Family structure
<i>Australian Temperament Project</i> (Renda et al. 2011)	Australia	1359 students	No: 7 * Parental substance use * Mother's age * Parental education * Parental occupation * Parental monitoring * Harsh discipline * Anti-social peer affiliations (20 covariates controlled for depression) ⁵⁵
<i>Cambridge Study in Delinquent Development</i> (Farrington & Ttofi, 2011)	England	(results on offending) 411 South London males followed-up from age 8–10 to age 48–50 (in the meta-analysis, results on conviction between ages 15–20)	No: 20 High daring; hyperactivity; high clumsiness; low non-verbal IQ; low verbal IQ; low attainment; high extraversion; high neuroticism; low popularity; low height; low weight; convicted parent; delinquent sibling; young mother; poor child rearing; disrupted family; low income; poor housing; low social class; large family size
<i>Christchurch Health and Development Study</i> (Gibb et al. 2011) ⁵⁶	New Zealand	979–985 students	No: 16 Gender; childhood conduct problems age 7–9; childhood sexual abuse age 0–16; deviant peer affiliations age 14; parental attachment age 15; childhood physical abuse age 0–16; IQ age 8/9; parental history of illicit drug use; family living standards age 0–10; childhood anxiety – – – withdrawal age 7–9; teacher-rated academic progress age 11–13; parental history of criminal offending; maternal age at participant's birth; maternal education
<i>Confident Kids Programme</i> (Berry and Hunt, 2009)	Australia	(results on depression) 24 male students of an intervention study (control group only)	No: 0

⁵⁵ For depression, the authors controlled for 20 covariates; Unpublished data provided via email communication with Jenny Renda, July 16, 2010. The results on victimization versus offending were also based on this email correspondence and 20 covariates were used for adjustment of effect sizes.

⁵⁶ For bullying victimization versus depression and offending, the authors have controlled for 14 covariates and not 16, as we have mistakenly indicated in the Ttofi et al., 2001 report in *JACPR*. For bullying perpetration versus the two outcomes, the authors have controlled for either 14 or 16 confounds depending on the age of the participants (results are shown separately for bullying in early childhood and adolescence). We assumed a total control of 16 confounds in the total summary effect size for bullying perpetration versus the outcomes.

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Danish Longitudinal Health Behaviour Study, Young Cohort</i> (Due et al., 2009)	Denmark	(results on depression) 847 children	No: 0
<i>Danish Longitudinal Retrospective Study</i> (Lund et al., 2008)	Denmark	(results on depression) 6097 males born in 1953 followed up in 2004 when aged 51	No: 2 -social class -parental mental illness (table 2; p. 113)
<i>Dunedin Longitudinal Study</i> (Moffitt et al. 2010)	New Zealand	(results on depression) ≈ 722 to 781 primary school-aged children (age 5–11)	No: 5 *Family SES * Childhood IQ (age 7–11) * Childhood neuroticism (age 5–11) * Childhood impulsivity (age 9–11) * Harsh parenting (age 7–9)
<i>Dutch Anti-Bullying Programme</i> (Fekkes et al., 2006)	the Netherlands	(results on depression) 1118 students aged 9–11 [the control group of a bullying-prevention programme]	No: 0
<i>Edinburgh Study of Youth Transitions and Crime</i> (Barker et al., 2008)	England	(results on delinquency) 3932 adolescents aged 14 to 16	No: 0 (trajectory analyses presented)
<i>Edinburgh Study of Youth Transitions and Crime</i> (Smith & Ecob, 2007)	England	(results on offending) 4300 adolescents from (annual) cohorts 2 to 6	No: 0 (latent class trajectories presented)
<i>Edinburgh Study of Youth Transitions and Crime</i> (McVie, 2010)	England	4299 students	No: 10 * gender * socio-economic status * impulsivity * peer delinquency scale * parental separation * parental supervision * parental conflict * parental punishment * neighbourhood deprivation * commitment to school
<i>E-Risk Longitudinal Study</i> (Arseneault, 2011; Shakoor et al., 2011)	England	(results on depression) Nationally representative UK birth cohort of 2232 same-sex twins	No: 5 Gender; SES status; maternal warmth; children's IQ; maltreatment

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Erlangen-Nuremberg Development and Prevention Study</i> (Lösel & Bender, 2011)	Germany	Bullying and victimization assessed in a sample of 557 children aged nine (wave 2) and followed up five years later (wave 3)	No: 5 Low family SES; family stressors; corporal punishment; conduct problems of child; emotional problems of child *confounds measured two years before bullying and victimization (wave 1)
<i>Erlangen-Nuremberg Longitudinal Study of School Bullying</i> (Bender and Lösel, 2011; Lösel & Bliesener, 2003; Lösel et al., 2008) *same results across the three reports	Germany	Of the 102 secondary school males aged 15 (in wave 2), 87 were contacted again in wave 3, 9 years later (no selective attrition from wave 2 to wave 3)	No: 3 * CBCL internalizing score * CBCL externalizing score * Comprehensive index of perceived family problems
<i>European TMR Network Project</i> (Singer, 2002)	England (and Spain for teacher data; not related to the current project)	(results on depression) 207 Spanish university students and 117 English university students (p. 123)	No: 7 * Personality factor (Neuroticism) *Personality factor (Openness) *Personality factor (conscientiousness) * Attachment type (preoccupied) *Friendship *Humour * Attachment type (fearful)
<i>Follow-Up Study in Canada</i> ⁵⁷ (Vaillancourt et al., 2011)	Canada	168 students (91 boys) aged 12, predominantly Caucasian (78%)	No: 0
<i>From a boy to a man Finnish Longitudinal Study; Part of Nationwide Finnish 1981 Birth Cohort Study</i> (Havisto et al. 2004)	Finland	(results on depression) 2348 male students aged 8, followed up when aged 18	No: 0
<i>From a boy to a man Finnish Longitudinal Study; Part of Nationwide Finnish 1981 Birth Cohort Study</i> (Klomek et al. 2008)	Finland	(results on depression) 2348 males born in 1981, followed-up at age 18 (bullying measured at age 8)	No: 1 * Depression at age 8
<i>From a boy to a man Finnish Longitudinal Study</i> (Sourander et al. 2007b)	Finland	(results on depression) 2540 male students aged 8, followed-up when aged 18–23	No: 0 (for other outcomes the authors control for: parental education level; and psychosis at age 8 based on parent/teacher reports, but not for depression; see page 400)

⁵⁷ In path analyses, Vaillancourt and colleagues (2011) offer standardized path coefficients for victimization versus depression after controlling for previous depression, but we are not interested in change analyses. Thus, these coefficients were not taken into account as adjusted effect sizes.

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Nationwide Finnish 1981 Birth Cohort Study</i> (Sourander et al. 2009)	Finland	(results on depression) 5038 males and females born in 1981	No: 1 * General Psychopathology at age 8
<i>From a boy to a man Finnish Longitudinal Study; Part of Nationwide Finnish 1981 Birth Cohort Study</i> (Sourander et al. 2006)	Finland	(results on offending) 2946 male students aged 8 followed up when aged 16–20	No: 4 * Living in other than 2-biological parent family * Mother's/father's educational level * Rutter A2 Total Scale * Rutter B2 Total Scale
<i>From a boy to a man Finnish Longitudinal Study; Part of Nationwide Finnish 1981 Birth Cohort Study</i> (Sourander et al. 2007a)	Finland	(results on offending) 2551 male students aged 8 followed up when aged 16–20	No: 2 *Parental education level * Child's psychosis (for different tables different covariates are used, showing different outcome measures on offending; only one confound in each data analyses)
<i>Nationwide Finnish 1981 Birth Cohort Study</i> (Sourander et al., 2011)	Finland	(results on offending) 5351 individuals (2639 girls) aged 23–26 (attrition rate of 11.1 of the initial sample)	No: 2 *Parental education level *Childhood total psychopathology at age 8 based on parent and teacher reports
<i>Gatehouse Project</i> (Bond et al., 2001)	Australia	(results on depression) 2680 students	No: 5 * Implementation group (experimental versus control) * Availability of attachments at baseline * Arguments with others at baseline * Sex * Family structure
<i>Health 2000 Project</i> (Pirkola et al. 2005)	Finland	(results on depression) 4706 Finnish individuals aged 30–64	No: 11 * sex, financial difficulties, parental unemployment, parental medical illness or injury, paternal problems with alcohol, maternal problems with alcohol, paternal mental health problems, maternal mental health problems, family discord, parental divorce, serious or long-term illness (table 4, p. 774)
<i>International Youth and Development Study</i> (Patton et al., 2008)	Bi-national study: U.S.A. and Australia (3 waves: 2002; 2003; and 2004)	(results on depression) 1701 female students in grades 7 and 9 (table 5; the paper presents data for male students and for grade 5, but these results are not relevant to the aims of our meta-analysis)	No: 11 * Pubertal stage change * Age * School grade level * State of origin * Previous level of depressive symptoms * Low family attachment * High family conflict * Low school connection * Self-blaming coping style * Poor emotional control * Low self-efficacy

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>International Youth Development Study</i> (Hemphill et al. 2011)	Australia	687 Year 7 students (long-term follow-up) and 701 Year 10 students (short-term follow-up)	No: 8 gender; student impulsivity; student attention deficits; antisocial peers; family history of antisocial behaviour; poor family management; family conflict; academic failure
<i>Japanese Longitudinal Study</i> ⁵⁸ (Nishino, 2010; Nishino et al. 2009, 2011)	Japan	(results on offending obtained via email) 532 children followed up for 5 months (results on depression) 330 (150 boys) 1st grade junior high school students followed-up for two years *differences in follow-up periods depending on the predictor; see text	(results on offending) No: 4 Negative attitude against school work; harsh parental discipline; low self-worth; deviant peers (results on depression) No: 4 Poor adjustment to school; harsh parental discipline; extreme peer orientation; interparental discord
<i>Jyvaskyla Longitudinal Study in Finland</i> (Pulkkinen & Tremblay, 1992)	Finland and Canada (comparisons of the two studies)	(results on offending) 369 Finnish children aged 8 followed up to the age of 26.	No: 0
<i>KiVa Anti-Bullying Programme</i> (Salmivalli, 2010)	Finland	(results on depression) ≈ 3000 to 4000 school children aged 9.5 [end of grade 3]	No: 0
<i>Longitudinal Retrospective Study at the Mood Disorders Unit Outpatient Depression Clinic</i> (Gladstone et al., 2006)	Australia	(results on depression) 222 adults assessed at an outpatient depression clinic	(applicable for anxiety, but not depression) No: 4 * Participants' age * Parental overcontrol * Behavioural inhibition * Childhood illness or disability
<i>Longitudinal Retrospective Study of Adult Twin Pairs</i> (Gladstone & Parker, 2006)	Australia	(results on depression) 576 randomly selected subjects from twin pairs (one of each pair)	No: 0
<i>Longitudinal Retrospective Study of American University Students</i> (Roth et al. 2002)	U.S.A	(results on depression) 514 university students (mean age 19.36, SD: 4.90)	No: 1 * Anxiety

⁵⁸ Most up-to-date results were provided by Dr Nishino via email communication; See relevant section in the text.

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Longitudinal Retrospective Study of English GBQ men</i> (Rivers, 1999, 2001; Rivers & Cowie, 2006)	England	A total of 235 bullied and non-bullied LGB individuals compared with 207 bullied and non-bullied heterosexual individuals (results on depression) 119 LGB bullied individuals compared with 116 LGB non-bullied individuals *Results in the meta-analysis based on the LGB individuals only; see text	No: 0
<i>Longitudinal Retrospective Study of Japanese University Students</i> (Matsui et al. 1996)	Japan	(results on depression) 134 male university students	No: 0
<i>Mater-University of Queensland Study of Pregnancy and its Outcomes</i> (McGee et al. (2011))	Australia	1806 children	No: 3 * Family poverty up to 14 * Physical punishment * CBCL measures of aggression and social/ thought disorders (composite score) at age 5
<i>Metropolitan Area Child Study; Young Cohort</i> (Henry et al., 2010)	U.S.A.	197 2nd and 3rd graders (≈ 8 years of age)	No: 4 covariates * family poverty * gender * parent involvement * impoverished community
<i>Metropolitan Area Child Study; Old Cohort</i> (Henry et al., 2010)	U.S.A.	259 5th and 6th graders (≈ 11 years of age)	No: 4 covariates * family poverty * gender * parent involvement * impoverished community
<i>Montreal Longitudinal Study</i> (Haapasalo et al., 2000; Tremblay & Haapasalo, 1998) *same data/information across the two reports	Montreal, Canada (the report by Pulkinnen & Tremblay 1992 presents data from the Jyvaskyla Longitudinal Study and the Montreal Longitudinal Study)	(results on offending) 1034 kindergarten boys aged 6.23 followed up to age 12 (Haapasalo et al., 2000: 148)	No: 1 Family Adversity

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Multimedia Violence Prevention Study</i> (Espelage et al., 2001)	U.S.A.	(results on depression) 558 6th, 7th and 8th graders	No: 0
<i>National Longitudinal Survey of Youth 1997</i> (Wong, 2009)	U.S.A.	(results on offending) 8833 school children aged 14,34 (p. 140), matched based on propensity score matching	No: 20 Sex; age; census region; race/ethnicity; household size; household income; non-english language at home; mother's age at birth; hours of child-care at home; participation in Head Start; number of schools attended; current school size; number of grades repeated; math exam score; physical environment risk index; enriching environment risk index; family routines index; parents' religiosity scale; substance use index; behavioural/emotional problems scale (see p. 136)
<i>Pittsburgh Youth Study</i> (White and Loeber, 2008)	U.S.A.	(results on delinquency) 421 boys in grades 2 through 5	No: 8 African-American; low SES; family adversity; neighbourhood disadvantage; aggression; poor academic achievement; special education; not liked by peers
<i>Pittsburgh Youth Study</i> (Farrinton et al., 2011a)	U.S.A.	503 boys of the youngest cohort originally assessed at age 6-7 in 1987-88	No: 10 Hyperactivity; low achievement; poor supervision; low reinforcement; poor communication; low involvement; delinquent peers; bad friends; low social class; poor housing
<i>Raising Health Children Project</i> (Kim et al. 2011)	U.S.A.	(results on offending) 957 children	No: 6 * Gender * Race /ethnicity * Low income status, grade 5 * Impulsivity, grade 6 * Poor family management, grade 5 * Antisocial peer association, grade 6
<i>SET Project</i> (Kimber et al., 2008 a & b)	Sweden	(results on depression) 761? children aged 11-16 (grades 5-9; Kimber et al., 2008, p. 932)	No: 3 Controlling for the intervention factor, SES and sex
<i>Seven Schools</i> (Ozdemir and Stattin, 2011)	Sweden	(results on depression) 508 children (53.2% girls) (results on offending) 880 students (unpublished data given by Kendrick via email)	No: 2 * Age * Gender (unadjusted effect sizes for offending)

Study /Authors	Country	Sample Size	Number of Covariates Controlled for
<i>Six-Month Follow-Up in Canada</i> (Shelley, 2009; Shelley & Craig, 2010) *same data across the two reports	Canada	(results on depression) 220 children	No: 0 (table 1: p. 26, chapter 2 from thesis)
<i>SNAP Under 12 Outreach Project</i> (Jiang et al. 2011)	Canada	(results on offending) 949 children (570 boys) aged 9.5	No: 5 * Gender EARLs Subscales: * Age at referral * Family subscale * Child subscale * Responsiveness Subscale
<i>Swedish Community Samples</i> (Olweus, 1993c, 1994b)	Sweden	(results on depression) 17 young men [identified as victims at grade 9] compared with 58 young men who were identified as non-victims at grade 9. All men followed-up to age 23	No: 0
<i>Swedish Community Samples</i> (Olweus, 2011) ⁵⁹	Sweden	(results on offending) 780 male students in grades 6 to 9	No: 0
<i>Swiss Federal Survey of Army Recruits of 1997</i> (Haas, 2001; Azzouzi and Killias, 2010; Staubli & Killias, 2011) *same results across three reports	Switzerland	(results on offending) 21,339 army recruiters aged 19–20 with questions on victimization at school between ages 6–11	No: 0
<i>z-prozo Longitudinal Study</i> (Averdijk et al. 2011)	Switzerland	(results on depression) Wave 2: 1320 (97% retention rate of the 1361 initial sample in wave 1) Wave 4: 1096 children	No: 11 Academic competence; competent problem solving; non-average height; weight; negative parenting; maternal depression; sex; siblings; SES; age of mother at birth; single parenthood

Note: Unless otherwise indicated, confounds are the same for the two outcome measures

⁵⁹ The Olweus 2011 report gives a detailed description of the study. In previous reports (i.e. Olweus, 1991; 1993 a, b, c; 1994 a, b; 1997) exactly the same ORs are presented but without the detailed description of the study's analytical procedures.

Tables 6 and 7 also present information that was used in the moderator analyses (see later), namely the mean age at which the predictor was measured, the mean age at which the outcome measure was taken, the number of covariates controlled for in the adjusted effect sizes, and whether the outcome measure was based on official data or self-reports⁶⁰. Information on moderators is presented separately for the two predictors (i.e. bullying perpetration and victimization), unless a study included data only on one of the two predictors.

The summary table 8 provides some further information (i.e. country of implementation, sample size, type of covariates controlled for). The sample size presented in table 8 does not necessarily correspond to that used in the meta-analyses. This is because specific data analyses in a report were sometimes based on a sub-sample of the dataset for methodological reasons. Excluded studies are not shown in this table. A final note about table 8 is that detailed information is presented for all included reports related to each longitudinal study⁶¹. To give an example, the SNAP Under 12 Outreach Project is represented in table 8 by one report while the Edinburgh Study of Youth Transitions and Crime is represented by three reports.

Following our exclusion criteria, some studies were excluded from the meta-analysis but not from the systematic review. In the section on inclusion and exclusion criteria (see above), we have already explained why various reports were excluded. For example, the Three-Year Follow-Up Study in Australia (Rigby, 1999, 2001) offers relevant data on the long-term effect of bullying victimization on anxiety, but not on depression. Questions on depression were excluded from the follow-up period upon explicit request from the schools (see Rigby, 1999: 98; Rigby, 2000: 320).

One of the reports on the Longitudinal Retrospective Study of English GBQ (Gay, Bisexual) men (Rivers, 2004) was excluded because comparisons were made between bullied LGB Gay, Bisexual) individuals with post-traumatic stress disorder and bullied LGB individuals with no post-traumatic stress disorder (both comparison groups were bullied; see Rivers 2004: table 3). Reports on comparing levels of depression between bullied and non-bullied LGB individuals were included in the meta-analysis. The same F value of 14.08 (based on the comparison between 119 bullied LGB individuals and 116 non-bullied LGB individuals) is reported in the Rivers (2001) paper, the Rivers and Cowie (2006) paper, and the Ph.D. thesis of Rivers (1999).

With regard to the E-Risk longitudinal study, data on depression are based on 23 items from the CBCL (Child Behaviour Checklist)

⁶⁰ This was applicable only in the case of offending/criminal behaviour.

⁶¹ We do not follow this rule when reports present exactly the same data (e.g. the Swedish Community Samples Study), in which case this is clearly indicated in the table.

and 27 items from the TRF (Teacher Report Form) on the withdrawn/anxious/depressed scales (Arseneault, 2011; Shakoor et al., 2011). A combined delinquency, aggression (and other externalizing problems) measure is also available in the same reports. Two other reports from the E-Risk Longitudinal Study (Bowes et al., 2009, 2010) were also excluded from all meta-analyses. Bowes et al. (2009) show ORs for bullying victimization versus internalizing and externalizing behaviour. However, internalizing/externalizing behaviour was measured at the baseline whilst bullying was measured in the follow-up period (Bowes et al., 2009: 550; see table 3). Subsequently, the paper was excluded since one of our inclusion criteria is that the predictor is measured before the outcome. This is also the main reason for excluding Bowes et al. (2010).

Finally, six of the reports related to the Finnish Cohort Longitudinal Study were excluded for various reasons. Kumpulainen and Rasanen (2000) show effect sizes for bullying (perpetration and victimization) at age 8 versus depression and internalizing/externalizing problems at age 15. This paper is based on the whole sample (boys and girls of the Nationwide Finnish 1981 Birth Cohort Study). In the tables with the effect size data, however, the authors (Kumpulainen & Rasanen, 2000: 1572) show the probability of being deviant when involved in bullying for the total parent and teacher scale (including internalizing/externalizing problems and depression as a total score) and do not show results separately for depression. In any case, a later paper by Sourander et al. (2009)⁶² shows results for a longer-term follow-up at age 24 for both the males and females.

In the paper by Kumpulainen and Roine (2002), bullying is included within other subscales and results are not presented separately for this predictor (see their table 1: 429). In the Klomek et al. (2009) study, results are shown for the association between bullying and suicides (in the adjusted effect sizes; see their table 2: 258) after controlling for depression. A paper by Kumpulainen et al. (2000) was excluded because the predictor (i.e. bullying) was part of another scale and results were not shown separately for bullying (see Kumpulainen et al., 2000: 6). A later paper by Kumpulainen et al. (2001) was also excluded because results on bullying versus depression were based on data within the same wave, making the report cross-sectional in character. The Sourander et al. (2000) paper was excluded as it did not provide data relevant to the aims of our review. Specifically, their tables 3 and 4 (pp. 877 – 878) show factors at age 8 (such as depression) predicting bullying perpetration and victimization at age 16. One of our inclusion criteria was that bullying (perpetration and victimization) had to precede the outcome measures.

⁶² Sourander et al. (2009) paper in *Archives of General Psychiatry*.

Moving on to included reports, clear rules were specified in advance regarding the way in which effect sizes would be combined within each report. These rules are explained in sections 2.5 and 2.6. As already mentioned, specific guidelines were also followed for combining effect sizes across reports relating to the same longitudinal study and these are presented briefly in section 2.7 and in more detail in Appendices 1 and 2.

2.5 Combining Effect Sizes within a Report: Bullying Perpetration/ Victimization versus Offending

We used Odds Ratios (ORs) as the measure of effect size. Where studies presented other statistics, these were converted into ORs (see the Technical Appendix in Ttofi et al., 2008). Within each manuscript more than one effect size could be reported. When choosing an appropriate effect size that would justify the inclusion of a report in the meta-analysis, the following rules were set:

1. Reports dealing with shoplifting, theft, vandalism/property damage, violent offending, arrest and police/court contact could be included in the meta-analysis.
2. Within a report, if different effect sizes were derived from official records of arrest or police/court contact, and from self-reports of shoplifting, theft, vandalism, or violent offending, these were combined into one effect size. However, if a general measure of offending as well as any of the specific offences were available within a report, then we chose to include the general measure in our meta-analysis. These rules avoided the inappropriate weighting of multiple effects.
3. If within a manuscript effect sizes were given separately for males and females, we combined the two measures. The same strategy was followed when separate measures were presented for two follow-up periods. It would have been ideal if we could have examined possible changes in the magnitude of the effect size within each study for different follow-up periods, but not many studies provided this information. We did, however, include the length of the follow-up period across studies in the moderator analyses. Very few studies presented data separately for males and females.
4. If for the same outcome measure different effect sizes were reported separately for each informant (e.g. teacher-rated, mother-rated, self-reported measures), but the manuscript also provided a combined measure across all informants, then we included the latter combined measure. We followed the same rule for the predictors (i.e. bullying perpetration and victimization), giving pref-

erence to a combined measure as opposed to a separate measure (e.g. we chose combined self- and peer-rated bullying rather than separate self- or peer-rated bullying).

5. In table 6, we list the reports from each longitudinal study. Under the name of each study, we indicate whether we have used a general measure of offending or a combined measure based on different criminal acts. The effect sizes are shown in table 9.

In Appendix 1, readers can find detailed descriptions of all includable reports relating to each longitudinal study. Each of these reports presents information which is relevant to one or both of our meta-analyses on offending outcomes. Only one report represents each longitudinal study in our meta-analyses (so that the overestimation of the effect size from dependent samples across reports is avoided) and the reasons for giving preference to one report over another is explained in the appendix.

Table 9. Effect Size Data for Offending

Author/Study Name/ Year	Offending T2 (Unadjusted ES)	Offending T2 (Adjusted ES)
<i>Australian Temperament Project (ATP)</i>	<i>19–20 years:</i>	<i>19–20 years:</i>
	B T1 vs.:	B T1 vs.:
	OR: 2.97	OR: 2.01
	(CI: 1.80 – 4.90)	(CI: 1.14 – 3.53)
	<i>23–24 years:</i>	<i>23–24 years:</i>
	B T1 vs.:	B T1 vs.:
	OR: 2.65	OR: 1.47
(CI: 1.14 – 6.12)	(CI: 0.58 – 3.77)	
<i>Cambridge Study in Delinquent Development (CSDD)</i>	<i>23–24 years:</i>	<i>23–24 years:</i>
	V T1 vs.:	V T1 vs.:
	<i>Property Damage:</i>	<i>Property Damage:</i>
	OR: 0.865	OR: 1.057
	(CI: 0.457 – 1.636)	(CI: 0.529 – 2.111)
	<i>Shoplifting:</i>	<i>Shoplifting:</i>
	OR: 0.9465	OR: 1.016
(CI: 0.457 – 1.958)	(CI: 0.410 – 2.517)	
<i>Cambridge Study in Delinquent Development (CSDD)</i>	B T1 vs.:	B T1 vs.:
	OR: 2.10	OR: 1.49
	(CI: 1.23 – 3.58)	(CI: 0.71 – 3.12)
	V T1 vs.:	V T1 vs.:
	---	---

Author/Study Name/ Year	Offending T2 (Unadjusted ES)	Offending T2 (Adjusted ES)
<i>Christchurch Health and Development Study (CHDS)</i>	<i>Arrest/Conviction:</i>	<i>Arrest/Conviction:</i>
	B _(childhood) T1 vs.:	B _(childhood) T1 vs.:
	OR: 4.8	OR: 2.5 (CI: 1.7 – 3.9)
	(CI: 3.4 – 6.7)	B _(adolescence) T1 vs.:
	B _(adolescence) T1 vs.:	OR: 1.4 (CI: 0.8 – 2.4)
	OR: 2.6	V _(adolescence) T1 vs.:
	(CI: 1.6 – 4.3)	OR: 1.2
	V _(adolescence) T1 vs.:	(CI: 0.6 – 2.5)
	OR: 2.3	<i>Property Offending:</i>
	(CI: 1.1 – 4.8)	B _(childhood) T1 vs.:
<i>Property Offending:</i>	OR: 1.3	
B _(childhood) T1 vs.:	(CI: 0.8 – 1.9)	
OR: 2.5	B _(adolescence) T1 vs.:	
(CI: 1.8 – 3.4)	OR: 0.8	
B _(adolescence) T1 vs.:	(CI: 0.5 – 1.5)	
OR: 1.6	V _(adolescence) T1 vs.:	
(CI: 1.0 – 2.6)	OR: 2.0	
V _(adolescence) T1 vs.:	(CI: 1.1 – 3.8)	
OR: 2.3		
(CI: 1.3 – 4.1)		
<i>Edinburgh Study of Youth Transitions and Crime (ESYTC)</i>	<i>Property Theft:</i>	<i>Property Theft:</i>
	B T1 vs.:	B T1 vs.:
	OR: 3.8	OR: 1.8
	(CI: 3.2 – 4.4)	(CI: 1.4 – 2.2)
	V T1 vs.:	V T1 vs.:
	OR: 1.5	OR: 1.4
(CI: 1.3 – 1.7)	(CI: 1.2 – 1.7)	
<i>Property Damage:</i>	<i>Property Damage:</i>	
B T1 vs.:	B T1 vs.:	
OR: 4.4	OR: 2.0	
(CI: 3.7 – 5.1)	(CI: 1.6 – 2.5)	
V T1 vs.:	V T1 vs.:	
OR: 1.1	OR: 0.8	
(CI: 0.9 – 1.2)	(CI: 0.7 – 0.9)	
<i>Erlangen-Nuremberg Development and Pre- vention Study (ENDPS)</i>	<i>Self-rated:</i>	<i>Self-rated:</i>
	B T1 vs.:	B T1 vs.:
	r = .24	r = .18
	[N = 557]	[N = 557]
	V T1 vs.:	V T1 vs.:
	r = .10	r = .03
	[N = 557]	[N = 557]
	<i>Mother-rated:</i>	<i>Mother-rated:</i>
	B T1 vs.:	B T1 vs.:
	r = .19	r = .11
	[N = 557]	[N = 557]
	V T1 vs.:	V T1 vs.:
r = .08	r = .03	
[N = 557]	[N = 557]	

Author/Study Name/ Year	Offending T2 (Unadjusted ES)	Offending T2 (Adjusted ES)
<i>Erlangen-Nuremberg Longitudinal Study of Bullying (ENLSB)</i>	BT1 vs.: r = .47 [N = 63] VT1 vs.: r = -.11 [N = 63]	BT1 vs.: r = .50 [N = 63] VT1 vs.: r = -.11 [N = 63]
<i>International Youth Deve- lopment Study (IYDS)</i>	<i>Year 7:</i> BT1 vs.: OR: 1.93 (CI: 1.10 – 3.40) VT1 vs.: OR: 0.94 (CI: 0.57 – 1.53) <i>Year 10:</i> BT1 vs.: OR: 3.50 (CI: 2.21 – 5.56) VT1 vs.: OR: 1.85 (CI: 1.14 – 2.98)	<i>Year 7:</i> BT1 vs.: OR: 1.21 (CI: 0.64 – 2.31) VT1 vs.: OR: 0.87 (CI: 0.51 – 1.50) <i>Year 10:</i> BT1 vs.: OR: 2.21 (CI: 1.27 – 3.85) VT1 vs.: OR: 1.63 (CI: 0.94 – 2.81)
<i>Japanese Longitudinal Study (JLS)</i>	<i>Shoplifting:</i> BT1 vs.: r = 0.25 [N = 532] VT1 vs.: r = 0.01 [N = 532] <i>Vehicle Theft:</i> BT1 vs.: r = 0.27 [N = 532] VT1 vs.: r = 0.01 [N = 532]	<i>Shoplifting:</i> BT1 vs.: r = 0.285 [N = 532] VT1 vs.: b = 0.019 [N = 532] <i>Vehicle Theft:</i> BT1 vs.: r = 0.249 [N = 532] VT1 vs.: b = 0.009 [N = 532]
<i>Jyvaskyla Longitudinal Study in Finland (JLSF)</i>	BT1 vs.: OR = 5.4166 (CI: 1.35 – 21.67) VT1 vs.: ---	BT1 vs.: --- VT1 vs.: ---
<i>Mater-University of Queensland Study of Pregnancy and its Out- comes (MUQSP)</i>	BT1 vs.: --- VT1 vs.: OR: 1.64 (CI: 0.99 – 2.72)	BT1 vs.: --- VT1 vs.: OR: 1.62 (CI: 0.98 – 2.69)
<i>Metropolitan Area Child Study; Young Cohort (MACS1)</i>	BT1 vs.: r = .17 [N = 197] VT1 vs.: r = .04 [N = 197]	BT1 vs.: r = .10 [N = 183] VT1 vs.: r = .00 [N = 183]

Author/Study Name/ Year	Offending T2 (Unadjusted ES)	Offending T2 (Adjusted ES)
<i>Metropolitan Area Child Study; Old Cohort (MACS2)</i>	B T1 vs.: r = .14 [N = 259] V T1 vs.: r = -.02 [N = 259]	B T1 vs.: r = .11 [N = 245] V T1 vs.: r = .00 [N = 245]
<i>Montreal Longitudinal Study (MLS)</i>	B T1 vs.: OR: 5.104 (CI: 2.39 – 10.897) V T1 vs.: ---	B T1 vs.: B = 1.63 (OR = 5.10) N = 311 Wald = 17.7 V T1 vs.: ---
<i>Nationwide 1981 Finnish Longitudinal Study (NFLS)</i>	<i>Males</i> <i>No bullying</i> N = 1127 1.2% > 5 crimes <i>Sometimes</i> <i>Bullying</i> N = 1215 4.5% > 5 crimes <i>Frequent Bullying</i> N = 213 10% > 5 crimes <i>No Victimization</i> N = 1098 3.1% > 5 crimes <i>Sometimes</i> <i>Victimization</i> N = 1227 3.6% > 5 crimes <i>Frequent Victimization</i> N = 244 5.7 % > 5 crimes <i>Females</i> <i>No Bullying</i> N = 2037 3.0% > 1 crimes <i>Sometimes and Frequent Bullying (Combined)</i> N = 612 4.6% > 1 crimes <i>No Victimization</i> N = 1098 3.1% > 1 crimes <i>Sometimes and Frequent Victimization (Combined)</i> N = 1227 3.8% > 1 crimes	<i>Males</i> <i>Sometimes</i> B T1 vs.: OR: 3.3 (CI: 1.8 – 6.2) V T1 vs.: OR: 0.7 (CI: 0.4 – 1.2) <i>Frequent</i> <i>Males</i> B T1 vs.: OR: 6.6 (CI: 2.8 – 15.3) V T1 vs.: OR: 0.6 (CI: 0.3 – 1.4) <i>Females</i> <i>Sometimes and Frequent (Combined)</i> B T1 vs.: OR: 1.3 (CI: 0.8 – 2.1) V T1 vs.: OR: 1.0 (CI: 0.6 – 1.6)

Author/Study Name/ Year	Offending T2 (Unadjusted ES)	Offending T2 (Adjusted ES)
<i>National Longitudinal Survey of Youth 1997 (NLSY)</i>	Arrest: B T1 vs.: --- V T1 vs.: OR = 1.52 (CI: 1.35 – 1.72) Theft: B T1 vs.: --- V T1 vs.: OR = 1.74 (CI: 1.55 – 1.94) Vandalism: B T1 vs.: --- V T1 vs.: OR = 1.88 (CI: 1.67 – 2.11) Other Property Crime: B T1 vs.: --- V T1 vs.: OR = 1.77 (CI: 1.52 – 2.06)	Arrest: B T1 vs.: --- V T1 vs.: OR = 1.19 (CI: 1.04 – 1.35) Theft: B T1 vs.: --- V T1 vs.: OR = 1.49 (CI: 1.32 – 1.68) Vandalism: B T1 vs.: --- V T1 vs.: OR = 1.56 (CI: 1.37 – 1.76) Other Property Crime: B T1 vs.: --- V T1 vs.: OR = 1.45 (CI: 1.23 – 1.70)
<i>Pittsburgh Youth Study (PYS)</i>	Boys: B T1 vs.: OR: 2.84 (CI: 1.85 – 4.36) V T1 vs.: --- Mothers: B T1 vs.: OR: 1.56 (CI: 1.28 – 1.91) V T1 vs.: ---	Boys: B T1 vs.: OR: 2.27 (CI: 1.45 – 3.53) V T1 vs.: --- Mothers: B T1 vs.: OR: 1.30 (CI: 1.02 – 1.64) V T1 vs.: ---
<i>Raising Health Children Project (RHCP)</i>	B T1 vs.: r = .16 [n = 957] V T1 vs.: ---	B T1 vs.: beta = .09 [n = 957] V T1 vs.: ---
<i>Seven Schools Longitudinal Study (SSLS)</i>	B T1 vs.: r = .22 [N = 870?] V T1 vs.: r = .09 [N = 870?]	B T1 vs.: --- V T1 vs.: ---
<i>SNAP Under 12 Out- reach Project (SU12OP)</i>	B T1 vs.: OR: 1.90 (CI: 1.11 – 3.26) V T1 vs.: ---	B T1 vs.: OR: 1.92 (CI: 1.08 – 3.41) V T1 vs.: ---
<i>Swedish Community Samples Study (SCSS)</i>	B T1 vs.: OR: 5.09 (CI: 2.948 – 8.835) V T1 vs.: ---	B T1 vs.: --- V T1 vs.: ---

Author/Study Name/ Year	Offending T2 (Unadjusted ES)	Offending T2 (Adjusted ES)
<i>Swiss Federal Survey of Army Recruits of 1997 (SFSAR)</i>	<i>Knifed Somebody:</i>	B T1 vs.: ---
	B T1 vs.: ---	V T1 vs.: ---
	V T1 vs.: OR: 2.975 (CI: 1.6046 – 5.5157)	
	<i>Strangled Somebody:</i>	
	B T1 vs.: ---	
	V T1 vs.:	
	OR: 1.898 (CI: 1.3554 – 2.6581)	
	<i>Shot with gun/stones:</i>	
	B T1 vs.: ---	
	V T1 vs.:	
	OR: 2.5833 (CI: 1.7015 – 3.9219)	
	<i>Shot with firearm:</i>	
B T1 vs.: ---		
V T1 vs.:		
OR: 2.8959 (CI: 1.6678 – 5.0282)		

2.6. Combining Effect Sizes within a Report: Bullying Perpetration/ Victimization versus Depression

Again, we used Odds Ratios (ORs) as the measure of effect size. Where studies presented other statistics, these were converted into ORs. Within each manuscript more than one effect size could be reported. When choosing an appropriate effect size that would justify inclusion of a report in the meta-analysis, the following rules were set:

1. Within a report, if different effect sizes were derived from self-, mother-, teacher-, peer-, and expert-rated depression, these were combined into one effect size (e.g. Lösel & Bender, 2011). We followed the same strategy for the predictors, i.e. bullying perpetration/victimization (e.g. self- and peer-rated bullying in the Salmivalli, 2010, report).
2. If a general measure of depression (based on a composite scale), as well as any of the specific items (or sub-scales), was available within a report, then we chose to include the general measure in our meta-analysis, unless we were restricted by the information presented in the study. For example, the Due et al. (2009) paper deals with depression based on the Beck Depression Inventory, but effect sizes of interest are shown based on a single item.
3. If the same informant filled in two different instruments on depression which were mutually exclusive (i.e. one not being a sub-

scale of the other), we have combined the relevant effect sizes (e.g. Bender & Lösel, 2011; Gladstone et al., 2006). We did not find any studies where participants filled in two different instruments on bullying perpetration/victimization.

4. If within a report effect sizes were given separately for males and females (e.g. Kaltiala-Heino et al., 2010; Nishino et al., 2011; Shelley, 2009) or for children at different ages (e.g. Hemphill et al., 2011)⁶³, we combined the two measures. The same combining strategy was followed when a separate measure was presented for two or more follow-up periods for the same individuals (e.g. Nishino et al., 2011; Ozdemir & Stattin, 2011). It would have been ideal if we could have examined possible changes in the magnitude of the effect size within each study for different follow-up periods, but not many studies provided this information. We did, however, include the length of the follow-up period in the moderator analyses.
5. If for the same outcome measure different effect sizes were reported separately for each informant, but the manuscript also provided a combined measure across all informants, then we chose to report the latter combined measure (e.g. a combined parent-teacher-child depression score for the z-proso longitudinal study by Averdijk et al., 2011). We followed the same rule for the predictors (i.e. bullying perpetration and victimization), giving preference to a combined measure as opposed to a separate measure (e.g. we chose the combined parent/teacher reports of bullying perpetration in middle childhood; see the relevant table 2 on page 85 of the Gibb et al., 2011, paper)⁶⁴.

In table 7, we list the reports from each longitudinal study. Under the name of each study, we indicate the mean age of the participants when the predictors (bullying perpetration/victimization) and depression were measured as well as the number of covariates (risk factors) that the authors controlled for when presenting the adjusted effect sizes. We also indicate whether the study is based on prospective or retrospective longitudinal data. If only one predictor is presented in table 7, this indicates that data were available only on this predictor. The actual effect sizes are shown in table 10.

In Appendix 2, readers can find detailed descriptions of all includable reports relating to each longitudinal study. Each of these reports presents information which is relevant to the aims of (one or both) of our meta-analyses on depression outcomes. Only one report represents each longitudinal study in our meta-analyses (so

⁶³ With the exception of those studies where younger and older children were based on different cohorts (e.g. Henry et al., 2010).

⁶⁴ For the E-Risk Study, a combined mother-child measure of bullying victimization was given by Louise Arseneault. However, the study was excluded for reasons explained before.

that an overestimation of the effect size from dependent samples across reports is avoided) and the reasons for giving preference to one report over another is explained in Appendix 2.

Table 10. Effect Size Data for Depression

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>Adolescent Mental Health Cohort Study (AMHC)</i>	Boys: B T1 vs.: OR: 4.3 (CI: 1.9 – 9.7) V T1 vs.: OR: 5.2 (CI: 2.4 – 11.1) Girls: B T1 vs.: OR: 2.2 (CI: 0.6 – 7.9) V T1 vs.: OR: 2.8 (CI: 1.1 – 7.3)	Boys: B T1 vs.: OR: 3.1 (CI: 1.2 – 7.7) V T1 vs.: OR: 4.6 (CI: 2.0 – 10.8) Girls: B T1 vs.: OR: 1.6 (CI: 0.4 – 7.4) V T1 vs.: OR: 1.8 (CI: 0.6 – 5.4)
<i>Australian Temperament Project (ATP)</i>	B T1 vs.: OR: 1.3039 (CI: 1.044 – 1.628) V T1 vs.: OR: 1.6313 (CI: 1.090 – 2.441)	B T1 vs.: OR: 1.331 (CI: 1.016 – 1.742) V T1 vs.: OR: 1.700 (CI: 1.056 – 2.736)
<i>Christchurch Health and Development Study (CHDS)</i>	B _(middle childhood) T1 vs.: OR: 1.6 (CI: 1.2 – 2.2) B _(adolescence) T1 vs.: OR: 1.7 (CI: 1.2 – 2.5) V T1 vs.: OR: 1.5 (CI: 0.8 – 2.6)	B _(middle childhood) T1 vs.: OR: 1.5 (CI: 1.1 – 2.2) B _(adolescence) T1 vs.: OR: 1.1 (CI: 0.8 – 1.9) V T1 vs.: OR: 1.2 (CI 0.6 – 2.2)
<i>Confident Kids Program (CKP)</i>	B T1 vs.: --- V T1 vs.: b = 0.187 [N = 24]	B T1 vs.: --- V T1 vs.: ---
<i>Danish Longitudinal Health Behaviour Study; Young Cohort (DLHBS)</i>	B T1 vs.: --- V T1 vs.: OR: 1.37 (CI: 0.83 – 2.26)	B T1 vs.: --- V T1 vs.: ---
<i>Danish Longitudinal Retrospective Study (DLRS)</i>	B T1 vs.: --- V T1 vs.: OR: 1.3365 (CI: 1.118 – 1.597)	B T1 vs.: --- V T1 vs.: OR: 1.24 (CI: 1.03 – 1.50)

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>Dunedin Longitudinal Study (DLS)</i>	<i>Diagnosed at Age 32 [N = 781]:</i> B T1 vs.: OR: 1.089 (CI: 0.920 – 1.289) V T1 vs.: ---	<i>Diagnosed at Age 32 [N = 754]</i> B T1 vs.: OR: 1.077 (CI: 0.872 – 1.330) V T1 vs.: ---
<i>Dutch Anti-bullying Programme (DAP)</i>	B T1 vs.: --- V T1 vs.: OR: 4.18 (CI: 1.87 – 9.36)	B T1 vs.: --- V T1 vs.: ---
<i>Edinburgh Study of Youth Transitions and Crime (ESYTC)</i>	B T1 vs.: OR: 1.7 (CI: 1.4 – 1.9) V T1 vs.: OR: 2.4 (CI: 2.1 – 2.7)	B T1 vs.: OR: 1.2 (CI: 1.0 – 1.4) V T1 vs.: OR: 2.2 (CI: 1.9 – 2.6)
<i>Erlangen-Nuremberg Development and Prevention Study (ENDPS)</i>	B T1 vs.: r = .11 [N = 557] V T1 vs.: r = .10 [N = 557]	B T1 vs.: r = .11 [N = 557] V T1 vs.: r = .08 [N = 557]
<i>Erlangen-Nuremberg Longitudinal Study of Bullying (ENLSB)</i>	<i>Depression</i> B T1 vs.: r = .26 [N = 57] V T1 vs.: r = .11 [N = 57] <i>Depressive PD</i> B T1 vs.: r = .31 [N = 57] V T1 vs.: r = .21 [N = 48]	<i>Depression</i> B T1 vs.: r = .32 [N = 57] V T1 vs.: r = .05 [N = 57] <i>Depressive PD</i> B T1 vs.: r = .28 [N = 52] V T1 vs.: r = .10 [N = 48]

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>European TMR Network Project (EuTMRNet)</i>	Primary: B T1 vs.: --- V T1 vs.: Verbal: r = .15 [N = 183] Indirect: r = .19 [N = 183] Secondary: B T1 vs.: --- Verbal: r = .16 [N = 183] *not included in the analy- ses; see notes	Primary & Secondary: (7 separate models) B T1 vs.: --- V T1 vs.: r = .10 [N = 280] r = .24 [N = 273] r = .23 [N = 268] r = .20 [N = 270] r = .27 [N = 273] r = .27 [N = 272] r = .24 [N = 281]
<i>Follow-Up Study in Canada (FUSC)</i>	B T1 vs.: --- V T1 vs.: r = .37 [N = 168]	B T1 vs.: --- V T1 vs.: ---
<i>'From a Boy to a Man' Finnish Longitudinal Study; Part of the Fin- nish Cohort Longitudinal Study (NFLS)</i>	B T1 vs.: OR: 4.4 (CI: 1.6 – 12.2) V T1 vs.: OR: 2.5 (CI: 0.7 – 8.5) B/V T1 vs.: OR: 6.9 (CI: 2.0 – 24.4)	B T1 vs.: OR: 3.3 (CI: 1.2 – 9.3) V T1 vs.: OR: 1.4 (CI: 0.4 – 5.2) B/V T1 vs.: OR: 3.8 (CI: 1.01 – 14.7)
<i>Gatehouse Project (GP)</i>	B T1 vs.: --- V T1 vs.: OR = 2.30 (CI: 1.2 – 4.3)	B T1 vs.: --- V T1 vs.: OR = 2.03 (CI: 1.14 – 3.64)
<i>Health 2000 Project (HEALTH2000)</i>	Males: B T1 vs.: --- V T1 vs.: OR: 2.96 (CI: 2.01 – 4.37) Females: B T1 vs.: --- V T1 vs.: OR: 2.23 (CI: 1.61 – 3.08)	Adjusted 1: Males: B T1 vs.: --- V T1 vs.: OR: 2.42 (CI: 1.61 – 3.62) Females: B T1 vs.: --- V T1 vs.: OR: 2.09 (CI: 1.51 – 2.90) Adjusted 2: Total Sample: B T1 vs.: --- V T1 vs.: OR: 2.20 (CI: 1.6 – 3.02)

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>International Youth Development Study (IYDS)</i>	<i>Year 7:</i>	<i>Year 7:</i>
	B T1 vs.:	B T1 vs.:
	OR: 1.10	OR: 0.97
	(CI: 0.71 – 1.72)	(CI: 0.58 – 1.63)
	V T1 vs.:	V T1 vs.:
	OR: 1.14	OR: 0.89
	(CI: 0.85 – 1.54)	(CI: 0.64 – 1.24)
	<i>Year 10:</i>	<i>Year 10:</i>
	B T1 vs.:	B T1 vs.:
	OR: 1.53	OR: 1.39
(CI: 0.99 – 2.36)	(CI: 0.85 – 2.27)	
V T1 vs.:	V T1 vs.:	
OR: 2.03	OR: 1.84	
(CI: 1.48 – 2.79)	(CI: 1.30 – 2.59)	
<i>Japanese Longitudinal Study (JLS)</i>	B T1 vs.:	B T1 vs.:
	r = 0.12	r = 0.045
	[N = 532]	[N = 532]
	V T1 vs.:	V T1 vs.:
	<i>Boys:</i>	<i>Boys:</i>
	Time 1 r = 0.26	Time 1 β = 0.15
	Time 2 r = 0.26	Time 2 β = 0.15
	Time 3 r = 0.15	Time 3 β = - 0.01
	[N = 150]	[N = 150]
	<i>Girls:</i>	<i>Girls:</i>
Time 1 r = 0.13	Time 1 β = 0.12	
Time 2 r = - 0.01	Time 2 β = - 0.02	
Time 3 r = 0.01	Time 3 β = - 0.04	
[N = 180]	N = 180]	
<i>Kiva Anti-bullying Programme (KIVA)</i>	<i>Self-rated:</i>	B T1 vs. ---
	B T1 vs.:	V T1 vs. ---
	r = .122	
	[N= 2275]	
	V T1 vs.:	
	r = .224	
	[N= 2274]	
	<i>Peer-rated:</i>	
	B T1 vs.:	
	r = .107	
[N= 2413]		
V T1 vs.:		
r = .127		
[N= 2413]		

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>Longitudinal Retrospective Study at the Mood Disorders Unit Outpatient Depression Clinic (MDUnit)</i>	B T1 vs.: --- V T1 vs.: <i>Based on BDI:</i> <i>Bullied:</i> Mean: 34.2 SD: 11.9 N = 54 <i>Not bullied:</i> Mean: 27.0 SD: 12.1 N = 151 <i>Based on HDRS:</i> <i>Bullied:</i> Mean: 15.8 SD: 6.6 N = 54 <i>Not bullied:</i> Mean: 15.2 SD: 7.4 N = 151	B T1 vs.: --- V T1 vs.: ---
<i>Longitudinal Retrospective Study of Adult Twin Pairs (LR-ATP)</i>	B T1 vs.: --- V T1 vs.: r = 0.096 [N = 576]	B T1 vs.: --- V T1 vs.: ---
<i>Longitudinal Retrospective Study of American University Students (LR-AUS)</i>	B T1 vs.: --- V T1 vs.: r = .38 [N = 514?]	B T1 vs.: --- V T1 vs.: r = .21 [N = 514?]
<i>Longitudinal Retrospective Study of English GBQ men (LR-GBQ)</i>	B T1 vs.: --- V T1 vs.: F = 14.08 N1 = 119 N2 = 116 d = .24 [SE = .131]	B T1 vs.: --- V T1 vs.: ---
<i>Longitudinal Retrospective Study of Japanese University Students (LR-JUS)</i>	B T1 vs.: --- V T1 vs.: r = .25 [N = 134]	B T1 vs.: --- V T1 vs.: ---
<i>Mater-University of Queensland Study of Pregnancy and its Outcomes (MUQSP)</i>	B T1 vs.: --- V T1 vs.: OR: 1.18 (CI: 0.62 – 2.25)	B T1 vs.: --- V T1 vs.: OR: 1.21 (CI: 0.63 – 2.31)

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>Metropolitan Area Child Study; Young Cohort (MACS1)</i>	B T1 vs.: $r = .04$ [N = 197] V T1 vs.: $r = .09$ [N = 197]	B T1 vs.: $r = .00$ [N = 183] V T1 vs.: $r = .07$ [N = 183]
<i>Metropolitan Area Child Study; Old Cohort (MACS2)</i>	B T1 vs.: $r = .14$ [N = 259] V T1 vs.: $r = .22$ [N = 259]	B T1 vs.: $r = .16$ [N = 245] V T1 vs.: $r = .20$ [N = 245]
<i>Multimedia Violence Prevention Study (MVPS)</i>	B T1 vs.: $r = .17$ [n = 516] V T1 vs.: ---	B T1 vs.: --- V T1 vs.: ---
<i>Pittsburgh Youth Study (PYS)</i>	Boys: B T1 vs.: --- V T1 vs.: OR: 1.358 (CI: 1.103 – 1.673) Mothers: B T1 vs.: --- V T1 vs.: OR: 2.10 (CI: 1.727 – 2.553)	Boys: B T1 vs.: --- V T1 vs.: OR: 1.271 (CI: 1.006 – 1.606) Mothers: B T1 vs.: --- V T1 vs.: OR: 2.026 (CI: 1.619 – 2.535)
<i>SET Project (SET)</i>	B T1 vs.: --- V T1 vs.: ---	B T1 vs.: --- V T1 vs.: $r = .108$ [N = 172]

Author/Study Name/ Year	Depression T2 (Unadjusted ES)	Depression T2 (Adjusted ES)
<i>Seven Schools Longitudinal Study (SSLS)</i>	<p><i>Follow-Up 1</i> B T1 vs.: $r = .112$ [N = 422] V T1 vs.: $r = .221$ [N = 422]</p> <p><i>Follow-Up 2</i> B T1 vs.: $r = .086$ [N = 417] V T1 vs.: $r = .234$ [N = 417]</p>	<p><i>Baseline</i> N (bullies) = 71 N (victims) = 76 N (b/v) = 109 N (neither) = 252</p> <p><i>Follow-Up 1</i> <i>Bullies</i> B (unstandardized): 0.044 <i>Victims</i> B (unstandardized): 0.128 <i>Bully-Victims</i> B (unstandardized): 0.269 <i>SD (depression F-U1):</i> .60925</p> <p><i>Follow-Up 2</i> <i>Bullies</i> B (unstandardized): 0.160 <i>Victims</i> B (unstandardized): 0.269 <i>Bully-Victims</i> B (unstandardized): 0.341 <i>SD (depression F-U2):</i> .61753</p>
<i>Six-Month Follow-Up in Canada (SMFUC)</i>	<p><i>Boys:</i> B T1 vs.: $r = .30$ V T1 vs.: $r = .15$ [N = 124]</p> <p><i>Girls:</i> B T1 vs.: $r = .26$ V T1 vs.: $r = .30$ [N = 113]</p>	<p><i>Boys:</i> B T1 vs.: --- V T1 vs.: ---</p> <p><i>Girls:</i> B T1 vs.: --- V T1 vs.: ---</p>
<i>Swedish Community Samples Study (SCSS)</i>	<p>B T1 vs.: --- V T1 vs.: --- $d = .87$ (CI: .70 – 1.03)</p>	<p>B T1 vs.: --- V T1 vs.: ---</p>
<i>z-proso Longitudinal Study (Z-PROSO)</i>	<p>B T1 vs.: --- V T1 vs.: $r = .166$ [N = 1320]</p>	<p>B T1 vs.: --- V T1 vs.: $\beta = .171$ [N = 1320]</p>

2.7 Combining effect sizes across reports relating to the same longitudinal study

For various longitudinal studies the researchers presented more than one (published or unpublished) report. In tables 6 and 7, the reader can see the number of reports/manuscripts relating to each longitudinal study. Some longitudinal studies were represented by one report (e.g., for depression, the Pittsburgh Youth Study) while others were represented in the systematic review (but not in the meta-analysis) by up to a maximum of ten reports (e.g., for depression, the Nationwide Finnish 1981 Birth Cohort Study).

When different reports relating to the same longitudinal study presented different effect sizes (because of differences, for example, in the sample size or in the follow-up period that the authors used), the combination of effect sizes across reports is not straightforward as these effect sizes are based on dependent samples. These dependencies must be taken into account, as ignoring them will result in standard errors that are too small, often by a large degree. In this case, the meta-analyst would need to identify independent sets for analysis (Wilson, 2010)⁶⁵. We did face this challenge in the current meta-analyses.

We have already explained above how we dealt with the issue of combining different effect sizes from different reports relating to the same longitudinal study. Obvious rules were set such as, for example, giving preference to the most up-to-date paper with the longest follow-up period. For example, we have chosen the Sourander et al. (2011) paper on criminality over the previous papers (Sourander et al., 2006, 2007a), which also had relevant data. We have also chosen the Sourander et al. (2011) paper because this is the only paper where results are presented for the females (i.e. results are based on the Nationwide Cohort Study and not the ‘From a boy to a man’ sub-study).

Of course, we had to choose the most appropriate paper from what was available, which is why detailed descriptions of all papers were given in Appendices 1 and 2. For example, the Sourander et al. (2009) paper on depression would have been the most ideal since it presents data from the Nationwide 1981 Birth Cohort Study (i.e. data on both males and females). However, as already explained, we could not include the effect sizes presented in that paper because of their statistical form. Subsequently, we have chosen the Klomek et al. (2008) paper instead, acknowledging the fact that results were presented only for one gender. More details about combining effect sizes are given in Appendices 1 and 2.

⁶⁵ Email communication with David B. Wilson (October 25, 2010).

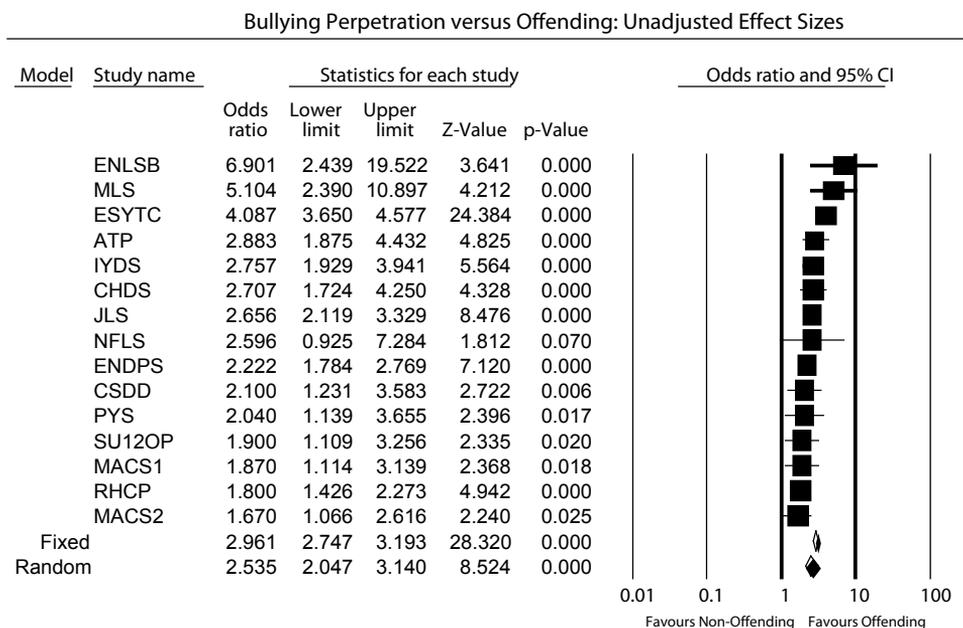
3. Bullying Perpetration versus Offending

3.1 Unadjusted and Adjusted Effect Sizes

Eighteen studies provided an effect size for bullying perpetration versus offending. For three of them (i.e. the Jyvaskyla Longitudinal Study, the Seven Schools Longitudinal Study and the Swedish Community Samples Study), only an unadjusted effect size was available. The summary unadjusted effect size across the 18 studies was $OR = 2.64$ (95% CI: 2.17 – 3.20; $z = 9.83$) for the random-effects model. We used the random-effects model since the heterogeneity test, Q , of 84.89 was highly significant at $p = .0001$. When the three studies with only unadjusted effect sizes were excluded, the summary effect size for the remaining 15 studies—for the random-effects model—was $OR = 2.54$ (95% CI: 2.05 – 3.14, $z = 8.52$). Again, there was significant variability in effect sizes across these studies ($Q = 76.03$, $p = .0001$). The summary effect size for each study was significant for all studies but one, as shown in the forest graph in figure 3.

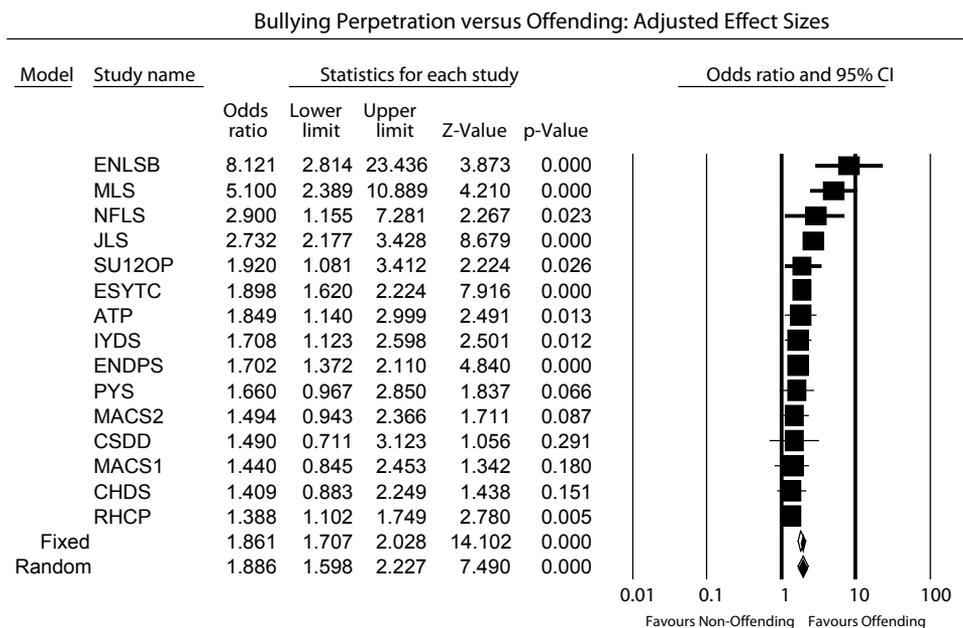
When controlling for covariates, the adjusted summary effect size was reduced to $OR = 1.89$, but this was still highly significant (95% CI: 1.60 – 2.23, $z = 7.49$). This OR indicates quite a strong relationship between bullying perpetration and later offending. For example, if a quarter of children were bullies and a quarter were offenders, this value of the OR would correspond to 34.5% of bullies becoming offenders, compared with 21.8% of non-bullies. Thus, being a bully increases the risk of being an offender (even after controlling for other childhood risk factors) by more than half. Figure 4 shows the forest graph for adjusted effect sizes. While all these effect sizes were in the expected direction, five were not statistically significant.

Figure 3. Unadjusted Effect Sizes for Bullying Perpetration versus Offending



Meta Analysis of Longitudinal Studies

Figure 4. Adjusted Effect Sizes for Bullying Perpetration versus Offending



Meta Analysis of Longitudinal Studies

3.2 Moderator Analyses

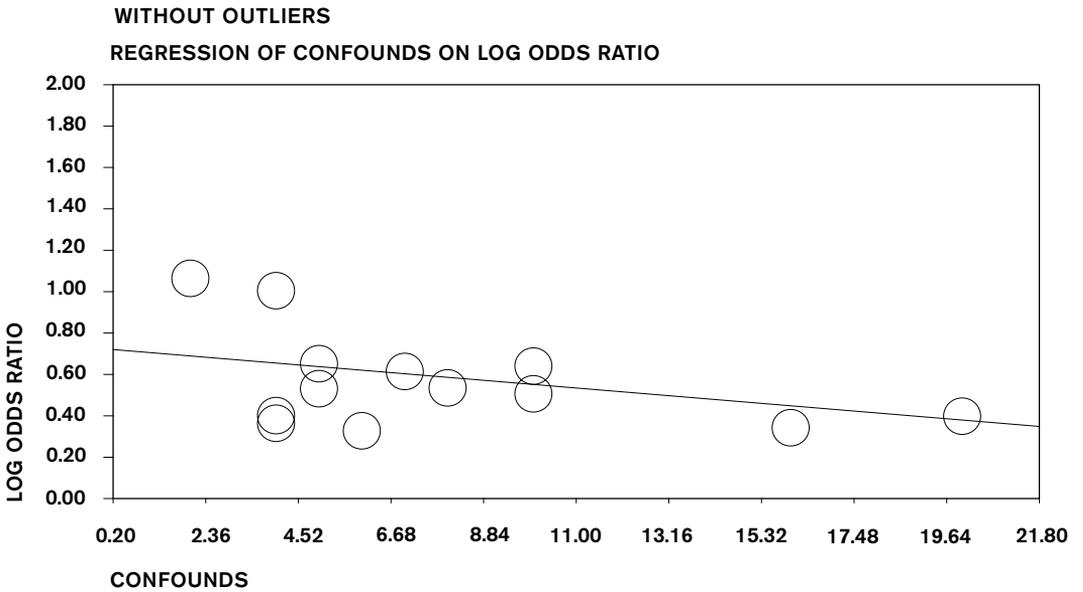
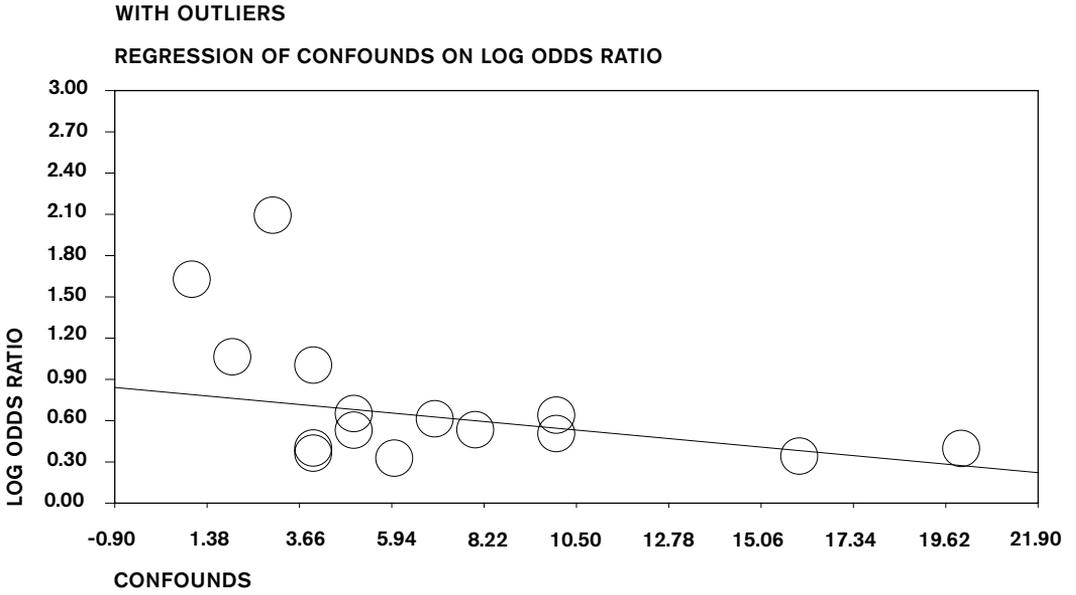
For the adjusted summary effect size, various moderators were investigated to explain the heterogeneity in effect sizes across studies, which was significant ($Q = 36.82, p = .001$). These included the number of covariates controlled for at baseline (range: 1 – 20; $M = 7.00$; $SD = 5.22$), the age at which school bullying was measured (range: 6.23 – 15.54; $M = 11.26$; $SD = 2.68$), the age of participants when outcome measures were taken (range: 10.00 – 24.64; $M = 17.10$; $SD = 4.91$) and the length of the follow-up period, measured in years (range: 0.42 – 16.50; $M = 5.84$; $SD = 4.56$).

The age at which bullying was measured was positively associated with the effect size, but the regression coefficient was not statistically significant ($B = .019, SE = .024, p = .428$). The length of the follow-up period was significantly negatively associated with the effect size ($B = -.027, SE = .012, p = .018$). As expected, the age of the study participants when outcome measures were taken was significantly negatively related to the effect size ($B = -.025, SE = .012, p = .039$). The above two negative relationships suggest that bullying perpetration has a stronger effect in the short-term. The relationship between the number of covariates controlled for and the effect size was in the expected negative direction and also significant ($B = -.027, SE = .013, p = .037$). Thus, the effect size decreased as the number of covariates controlled for increased.

Figure 5 shows that effect sizes were linearly (negatively) related to the number of covariates controlled for in all studies except two. When the two outliers were removed, the p-value for the unbiased regression coefficient was not significant ($B = -.017, SE = .013, p = .199, \text{intercept} = .724$). Despite the non-significant p-value, there remained a tendency for the effect size to decrease as the number of confounds increased.

Other moderators that may explain variability in effect sizes include the type of longitudinal studies (i.e. prospective versus retrospective) and the way in which the outcomes were measured (i.e. official data versus self-reports). In table 6, the reader can obtain information about these moderators. Only three studies out of fifteen presented outcome measures based on official records of offending (the Cambridge Study in Delinquent Development; the Nationwide Finnish 1981 Birth Cohort Study; and the Jyväskylä Longitudinal Study), making a moderator analysis inappropriate (due to uneven study numbers). Finally, only one study (i.e. the Swiss Federal Survey of Army Recruits of 1997) presented results based on a retrospective measure of bullying victimization, so any analyses comparing prospective and retrospective studies would be meaningless.

Figure 5. Relation between the Effect Size (Bullying Perpetration versus Offending) and the Number of Covariates (Confounds)



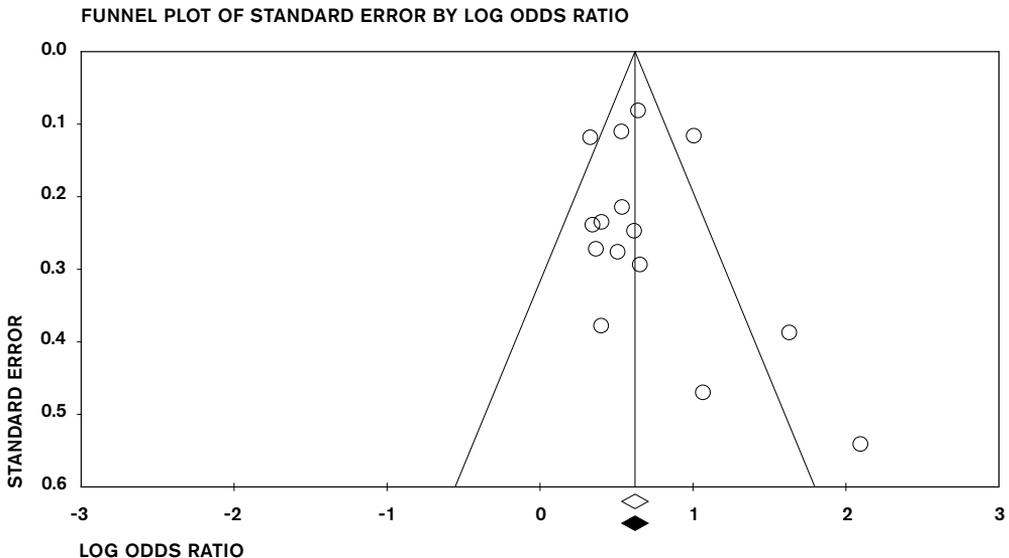
3.3 Publication Bias Analyses

If the studies included in a meta-analysis are a biased sample of all relevant studies, then the average effect size will reflect this bias (Borenstein et al., 2009: 277). As already shown in our searching strategies, we took every precaution to ensure that all eligible studies would be included in our meta-analyses. Nevertheless, in order to further increase the validity of our meta-analyses, we carried out a number of publication bias analyses.

Firstly, we used the Duval and Tweedie's Trim-and-Fill procedure. This technique displays the differences in effect sizes that could be attributable to bias by imputing effect sizes until the error distribution more closely approximates normality, offering the best *estimate* of the unbiased effect size (Borenstein et al., 2009, p. 286). As shown in figure 6, no imputed effect sizes are presented on the relevant funnel plot (they would have been presented as solid black dots). The imputed summary effect size (represented by a solid black diamond) has not shifted at all.

Indeed, under the fixed effect model the point estimate and 95% confidence interval for the combined studies is 1.86054 (95% CI: 1.70672, 2.02823). Using Trim and Fill these values remained unchanged. Under the random effects model the point estimate and 95% confidence interval for the combined studies is 1.88647 (95% CI: 1.59778, 2.22732). Using Trim and Fill these values were again unchanged.

Figure 6. Funnel Plot of Standard Error by Log Odds Ratio with Actual and Imputed Summary Effect Size (Bullying Perpetration versus Offending)



Furthermore, we conducted Rosenthal's Fail-Safe N test (Rosenthal, 1979). One concern of publication bias is that some non-significant studies are missing from a given analysis and that these studies, if included, would nullify the observed effect. Rosenthal suggested that, rather than simply speculate about the impact of the missing studies, we compute the number of non-significant studies that would be required to nullify the effect. If this number is small, then there is reason for concern because some non-significant studies may have been never communicated to the scientific community (e.g. due to 'publication bias'). However, if this number is large, we can be confident that the treatment effect, while possibly inflated by the exclusion of some studies, is nevertheless not zero.

This meta-analysis incorporates data from 15 studies, which yield a z-value of 12.69397 and corresponding 2-tailed p-value of 0.000001. The fail-safe N is 615. This means that we would need to locate and include 615 'null' studies in order for the combined 2-tailed p-value to be less than 0.05. Put another way, 41 missing studies for every observed study would be needed to nullify our effect. It is impossible for us to have missed such a large number of studies, or for such a large number of studies to have been carried out but not published.

The classic case of publication bias is the case depicted by the funnel plot. Large studies tend to be included in the analysis regardless of their treatment effect, whereas small studies are more likely to be included when they show a relatively large treatment effect. Under these circumstances there will be an inverse correlation between study size and effect size. Begg and Mazumdar (1994) suggested that this correlation can serve as a test for publication bias. Concretely, they suggest that we compute the rank order correlation (Kendall's tau b) between the treatment effect and the standard error (which is driven primarily by the sample size)⁶⁶.

In this case Kendall's tau b (corrected for ties, if any) is 0.24762, with a 1-tailed p-value (recommended) of 0.09911 or a 2-tailed p-value of 0.19821 (based on a continuity-corrected normal approximation). In conclusion, there is again no evidence of publication bias in this meta-analysis.

⁶⁶ This approach is limited in some important ways. A significant correlation suggests that bias exists but does not directly address the implications of this bias. Alternatively, it is possible that the smaller studies were of better quality and that there is no bias. Conversely, a non-significant correlation may be due to low statistical power, and cannot be taken as evidence that bias is absent.

4. Bullying Victimization versus Depression

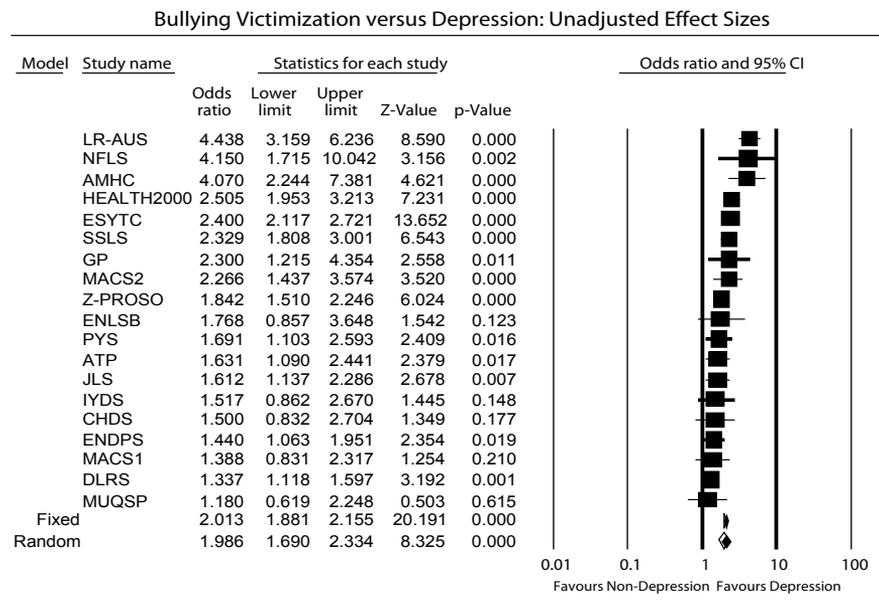
4.1 Unadjusted and Adjusted Effect Sizes

Thirty studies provided an effect size for bullying victimization versus depression. Only an unadjusted effect size was available for 11 of them (i.e. the Confident Kids Programme; the Danish Longitudinal Health Behaviour Study; the Dutch Anti-Bullying Programme; the Follow-Up Study in Canada; the Kiva Anti-Bullying Programme; the Longitudinal Retrospective Study at the Mood Disorders Unit; the Longitudinal Retrospective Study of Adult Twin Pairs; the Longitudinal Retrospective Study of English GBQ men; the Longitudinal Retrospective Study of Japanese University Students; the Six-Month Follow-Up in Canada; the Swedish Community Samples Study). Of these 11 studies, 5 were based on a retrospective measure of victimization, 3 were related to a follow-up intervention programme, and 3 were prospective studies where individuals were followed-up for various reasons but not because of an intervention study.

The summary effect size across the 30 studies was $OR = 2.08$ (95% CI: 1.80 – 2.41; $z = 9.84$) for the random-effects model. We used the random-effects model since the heterogeneity test, Q , of 128.87 was highly significant at $p = .0001$. When the 11 studies with only unadjusted effect sizes were excluded, the summary effect size for the remaining 19 studies—for the random-effects model—was $OR = 1.99$ (95% CI: 1.69 – 2.33, $z = 8.33$). Again, there was significant variability in effect sizes across these studies ($Q = 76.60$, $p = .0001$). The effect sizes for the majority of studies were significant, as shown in the forest graph in figure 7.

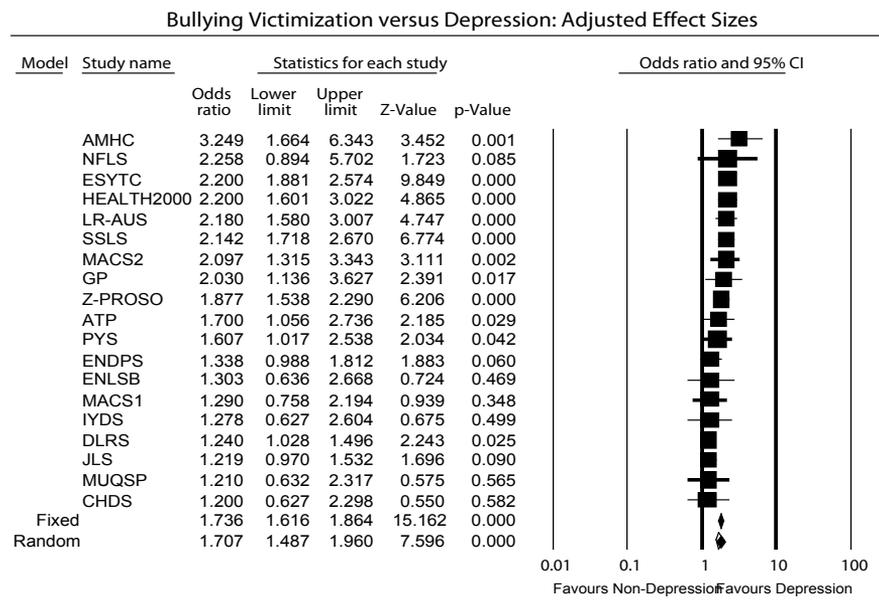
After controlling for covariates, the adjusted summary effect size

Figure 7. Unadjusted Effect Sizes for Bullying Victimization versus Depression



Meta Analysis of Longitudinal Studies

Figure 8. Adjusted Effect Sizes for Bullying Victimization versus Depression



Meta Analysis of Longitudinal Studies

was reduced to $OR = 1.71$, but this was still highly significant (95% CI: 1.49 – 1.96, $z = 7.60$) and with marked precision as shown by the narrow confidence intervals⁶⁷. This OR indicates quite a strong relationship between bullying victimization and later depression. For example, if a quarter of children were victims and a quarter were depressed, this value of the OR would correspond to 33.0% of victims becoming depressed, compared with 22.3% of non-victims. Thus, being a victim increases the risk of being depressed (even after controlling for other childhood risk factor) by about half. Figure 8 shows the forest graph for adjusted effect sizes. As with the previous forest plot, all the effect sizes were in the expected direction.

4.2 Moderator Analyses

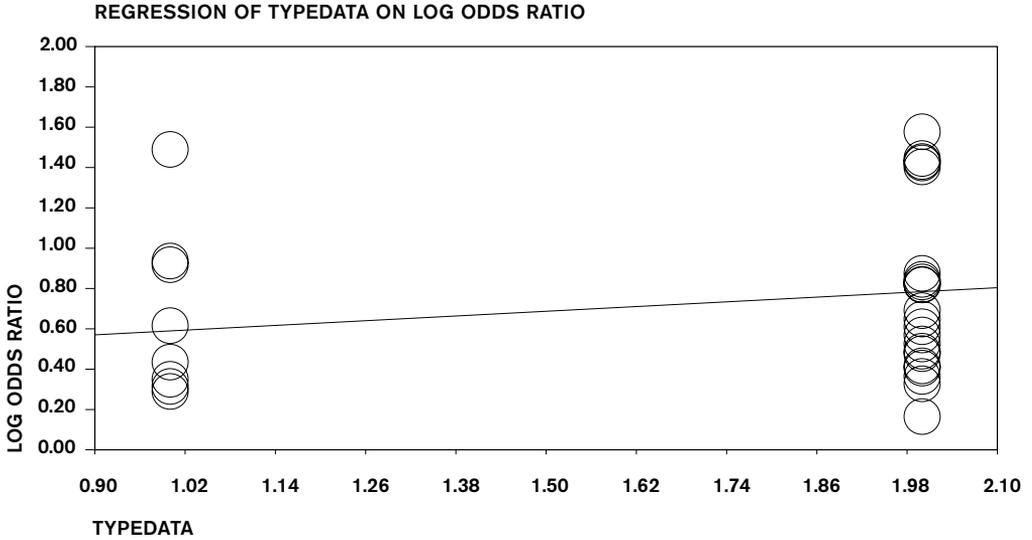
For the adjusted summary effect size, various moderators were investigated to explain the heterogeneity in effect sizes across studies, which was significant ($Q = 50.88$, $p = .0001$). These included the number of covariates controlled for at baseline (range: 1 – 20; $M = 6.42$; $SD = 5.06$), the age at which school bullying was measured (range: 8.00 – 18.00; $M = 12.32$; $SD = 2.74$), the age of participants when depressive symptoms were assessed (range: 10.00 – 47.00; $M = 19.45$; $SD = 9.64$) and the length of the follow-up period, measured in years (range: 1.00 – 36.00; $M = 7.13$; $SD = 8.79$). Because $I^2 = 64.62\%$, most of the between-study variation reflects real differences rather than random error. Therefore, fixed effects meta-regressions were used.

The age at which bullying victimization was measured was significantly negatively associated with the effect size ($B = -.028$, $SE = .012$, $p = .026$) and so was the age at which the outcome measure was taken ($B = -.007$, $SE = .003$, $p = .026$). As expected, the length of the follow-up period was significantly negatively related to the effect size ($B = -.007$, $SE = .004$, $p = .055$), suggesting that the deleterious effects of bullying victimization decrease as time goes by. Surprisingly, the relationship between the number of covariates controlled for and the effect size was not in the expected negative direction and it was also significant ($B = .020$, $SE = .008$, $p = .017$). This possibly reflects the many uncontrolled variables in this relationship.

As mentioned, nineteen studies provided both unadjusted and adjusted effect sizes for bullying victimization versus depression. Of these 19 studies, only three were based on a retrospective measure of bullying victimization. Therefore, we decided to conduct a

⁶⁷ Two studies (i.e. the European TMR Network Project and the SET Project) provided only adjusted effect sizes. Across the 21 studies, the adjusted summary effect size was $OR = 1.74$ (95% CI: 1.53 – 1.98; $z = 8.33$); $Q = 59.76$, $p = .0001$

Figure 9. Relationship between the Effect Size and the Types of Studies (Bullying Victimization versus Depression)



Note: 1 = retrospective, 2 = prospective.

moderator analysis on the type of research design for the 30 studies with unadjusted effect sizes. Studies with a longitudinal retrospective measure of bullying victimization were coded ‘1’, and prospective studies were coded ‘2’. The type of study was significantly positively related to the effect size ($B = 0.195$, $SE = 0.068$, $p = 0.004$). Longitudinal retrospective studies yielded a smaller effect size (see figure 9). The summary effect size for the eight longitudinal retrospective studies was $OR = 1.95$ (95% CI: 1.39 – 2.73; $z = 3.90$; $Q = 49.70$, $p = .0001$). The summary effect size for the twenty-two longitudinal prospective studies was $OR = 2.14$ (95% CI: 1.83 – 2.50; $z = 9.43$; $Q = 70.99$, $p = .0001$).

We may have expected larger effects in retrospective studies as they are more vulnerable to bias (attributable to problems of recollection of events by study participants). However, the retrospective studies in our analysis tended to have longer follow-up periods and we have found a negative relationship between the length of the follow-up period and effect size. This may explain the smaller effect sizes found in the retrospective studies.

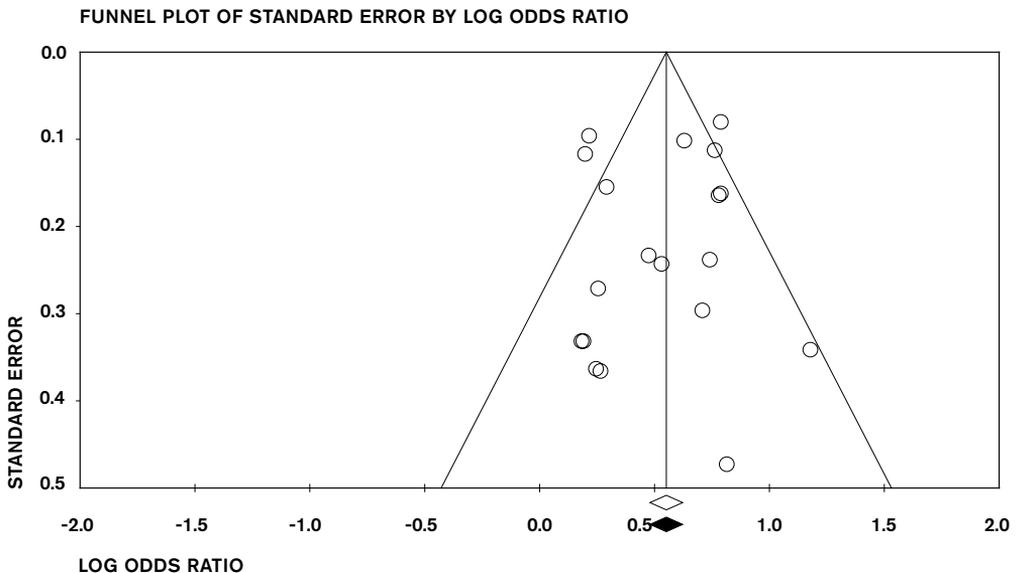
4.3 Publication Bias Analyses

Again, we carried out various sensitivity analyses. These were

based on the adjusted effect sizes for the comparable studies (i.e. the studies with both unadjusted and adjusted effect sizes). Firstly, we used the Duval and Tweedie's Trim-and-Fill procedure and the relevant plot suggests no bias (see figure 10). Under the fixed effect model the point estimate and 95% confidence interval for the combined studies is 1.73590 (95% CI: 1.61644, 1.86418). Using Trim and Fill these values remained unchanged. Under the random effects model the point estimate and 95% confidence interval for the combined studies is 1.70702 (95% CI: 1.48701, 1.95958). Using Trim and Fill these values were unchanged.

We have also conducted Rosenthal's Fail-Safe N test. The fail-safe N was 816. This means that one would need to locate and include 816 'null' studies in order for the combined 2-tailed p-value to be less than 0.050. This would require 43 missing studies for every observed study for the effect to be nullified, which is highly unrealistic. Finally, we conducted the Begg and Mazumdar rank correlation test. Kendall's tau b (corrected for ties, if any) is -0.08187, with a 1-tailed p-value (recommended) of 0.31214 or a 2-tailed p-value of 0.62428 (based on continuity-corrected normal approximation). Therefore, there is no evidence of publication bias.

Figure 10. Funnel Plot of Standard Error by Log Odds Ratio with Actual and Imputed Summary Effect Size (Bullying Victimization versus Depression)



5. Bullying Victimization versus Offending

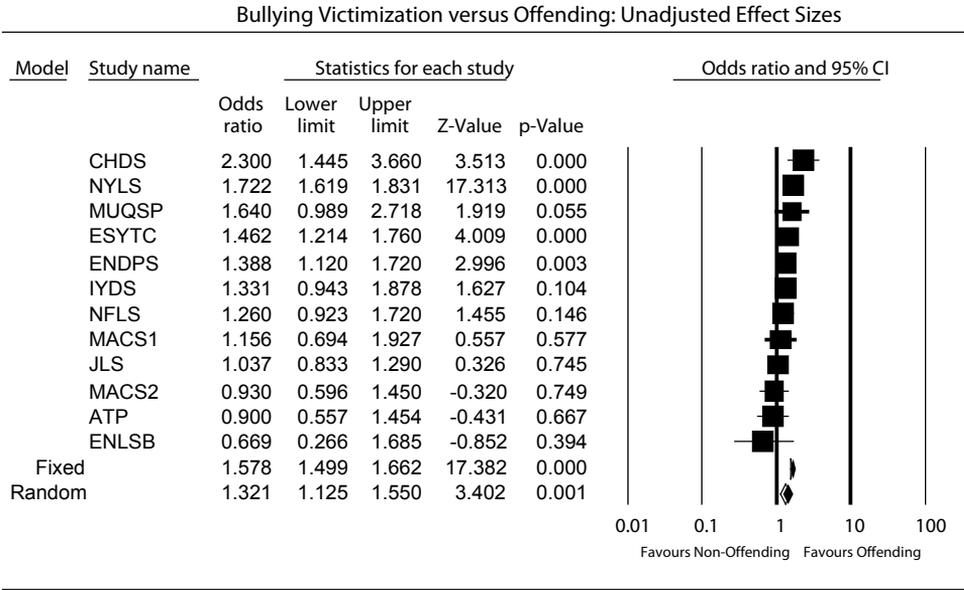
5.1 Unadjusted and Adjusted Effect Sizes

Fourteen studies provided an effect size for bullying victimization versus offending. For two of them (i.e. the Seven Schools Longitudinal Study and the Swiss Federal Survey of Army Recruits of 1997) only an unadjusted effect size was available.

The summary effect size across the 14 studies was $OR = 1.40$ (95% CI: 1.21 – 1.62; $z = 4.46$) for the random-effects model. We used the random-effects model since the heterogeneity test, Q , of 57.97 was highly significant at $p = .0001$. When the two studies with only unadjusted effect sizes were excluded, the summary effect size for the remaining 12 studies—for the random-effects model—was $OR = 1.32$ (95% CI: 1.13 – 1.55, $z = 3.40$). Again, there was significant variability in effect sizes across these studies ($Q = 44.86$, $p = .0001$). The summary effect size for a number of studies was not significant and, interestingly, in the opposite direction (i.e. supporting non-offending), as shown in the forest graph in figure 11.

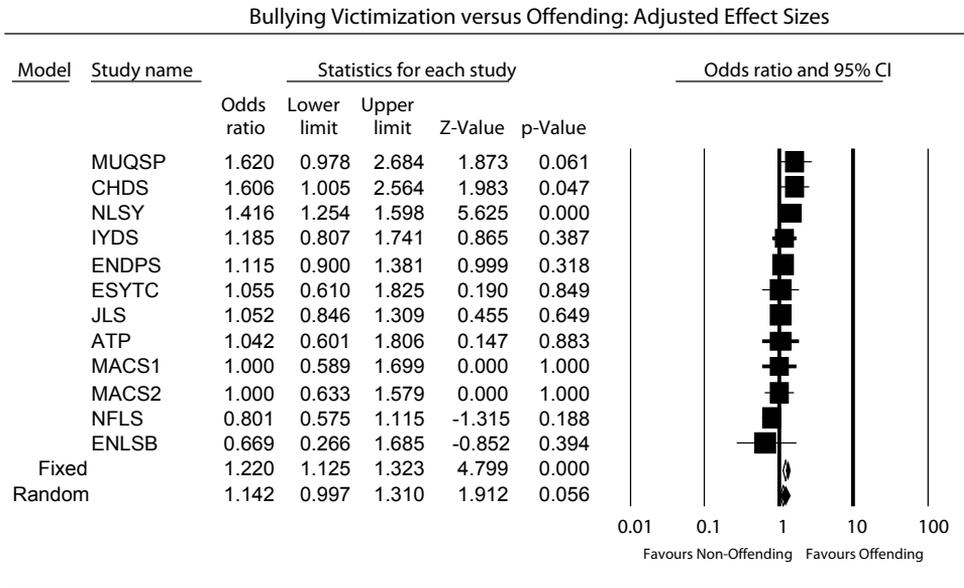
When controlling for covariates, the adjusted summary effect size was reduced to $OR = 1.14$. That was very nearly significant (CI: 0.997– 1.310, $z = 1.91$). The random effects model was used because the effect sizes were significantly heterogeneous ($Q = 20.51$, $p = .04$). This OR indicates only a weak relationship between bullying victimization and later offending. For example, if a quarter of children were victims and a quarter were offenders, this value of the OR would correspond to 26.9% of victims becoming offenders, compared with 24.4% of non-victims. Thus, being a victim increases the risk of being an offender (after controlling for other childhood risk factors) by only about 10%. Figure 12 shows the forest graph for adjusted effect sizes.

Figure 11. Bullying Victimization versus Offending: Unadjusted Effect Sizes



Meta Analysis of Longitudinal Studies

Figure 12. Bullying Victimization versus Offending: Adjusted Effect Sizes



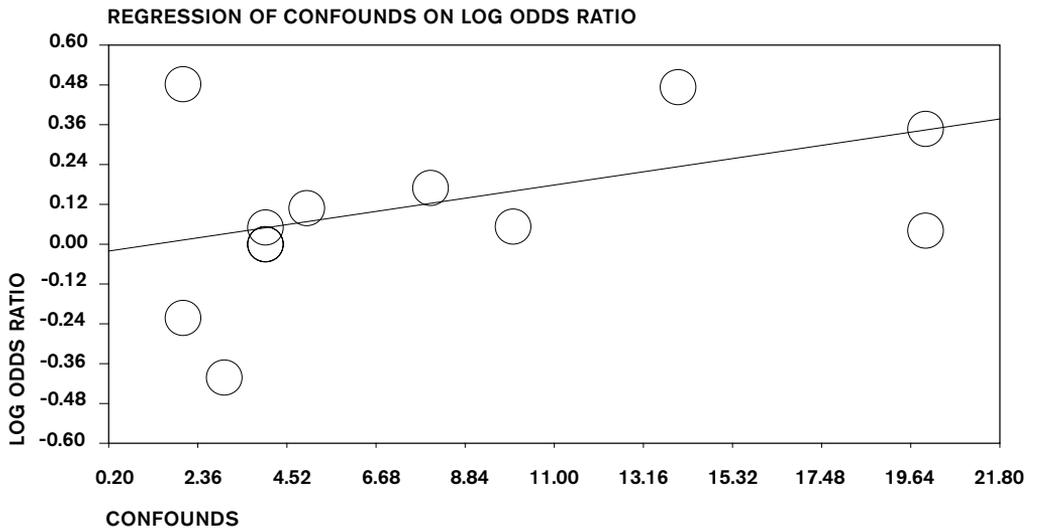
Meta Analysis of Longitudinal Studies

5.2 Moderator Analyses

Various moderators were investigated to explain the heterogeneity in effect sizes across studies. These included the number of covariates controlled for at baseline (range: 2 – 20; $M = 8.00$; $SD = 6.62$), the age at which school victimization was measured (range: 8 – 15.54; $M = 12.08$; $SD = 2.55$), the age of participants when outcome measures were taken (range: 10.00 – 24.64; $M = 17.62$; $SD = 5.33$) and the length of the follow-up period, measured in years (range: 0.42 – 16.50; $M = 5.55$; $SD = 4.85$).

The age at which victimization was measured was significantly positively associated with the effect size ($B = .050$, $SE = .023$, $p = .032$), while the length of the follow-up period was significantly negatively related with the effect size ($B = -.021$, $SE = .011$, $p = .052$). The age of the study participants when outcome measures were taken ($B = -.012$, $SE = .012$, $p = .295$) was also in the expected negative direction but the coefficient was not statistically significant. The relationship between the number of covariates controlled for and the effect size was not in the expected direction and was significant ($B = .018$, $SE = .005$, $p = .0006$). This again may reflect the influence of uncontrolled variables.

Figure 13. Relationship between the Effect Size (Bullying Victimization versus Offending) and the Number of Covariates (Confounds)



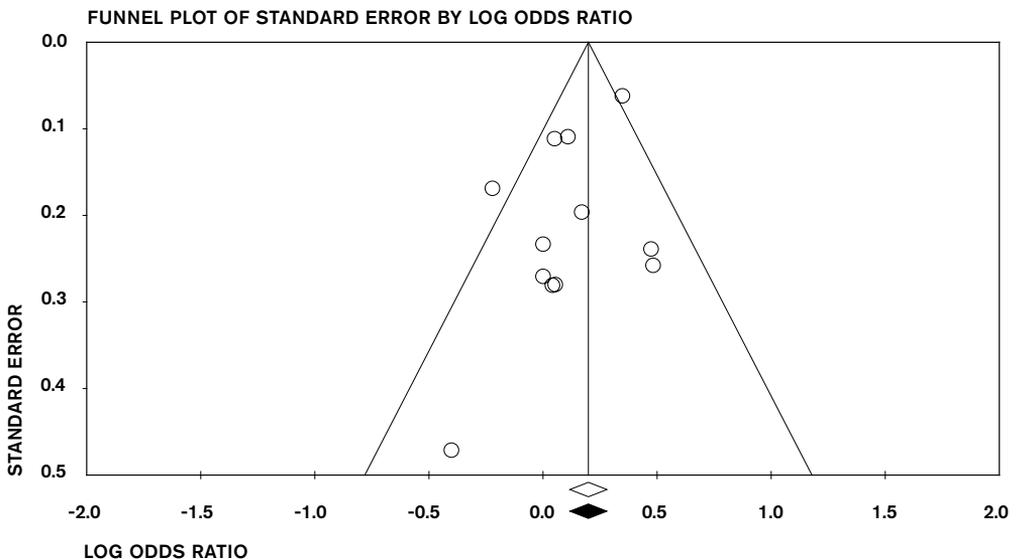
5.3 Publication Bias Analyses

Firstly, we used the Duval and Tweedie's Trim-and-Fill procedure. The funnel plot in figure 14 shows that the imputed summary effect size (represented by a solid black diamond) did not shift at all. Under the fixed effect model the point estimate and 95% confidence interval for the combined studies is 1.22010 (95% CI: 1.12490, 1.32337). Using Trim and Fill these values are unchanged. Under the random effects model the point estimate and 95% confidence interval for the combined studies is 1.14247 (95% CI: 0.99665, 1.30964). Using Trim and Fill these values are unchanged.

Furthermore, we conducted Rosenthal's Fail-Safe N test. This meta-analysis incorporates data from 12 studies, which yield a z-value of 2.87789 and corresponding 2-tailed p-value of 0.004. The fail-safe N is 14. This means that we would need to locate and include 14 'null' studies in order for the combined 2-tailed p-value to exceed 0.05. It is not plausible that we missed more studies than we retrieved in our thorough search process.

Finally, we conducted the Begg and Mazumdar rank correlation test. In this meta-analysis Kendall's tau b (corrected for ties, if any) was -0.01515, with a 1-tailed p-value (recommended) of 0.47266 or a 2-tailed p-value of 0.94533 (based on continuity-corrected normal approximation). Again, there was no evidence of publication bias.

Figure 14. Funnel Plot of Standard Error by Log Odds Ratio with Actual and Imputed Summary Effect Size (Bullying Victimization versus Offending)



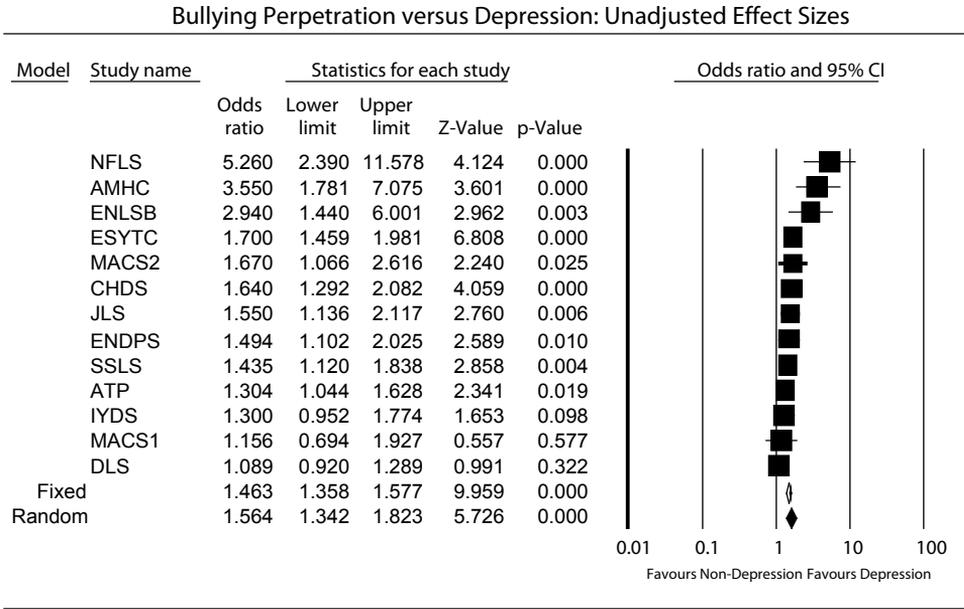
6. Bullying Perpetration versus Depression

6.1 Unadjusted and Adjusted Effect Sizes

Sixteen studies provided an effect size for bullying perpetration versus depression. For three of them (i.e. the Kiva Anti-Bullying Programme, the Multimedia Violence Prevention Study and the Six-Month Follow-Up in Canada), only an unadjusted effect size was available. The summary effect size across the 16 studies was $OR = 1.61$ (95% CI: 1.42 – 1.82; $z = 7.38$) for the random-effects model, with a heterogeneity test, Q , of 48.64 that is highly significant at $p = .0001$. When the three studies with only unadjusted effect sizes were excluded, the summary effect size for the remaining 13 studies—for the random-effects model—was very similar: $OR = 1.56$ (95% CI: 1.34 – 1.82, $z = 5.73$). Again, there was significant variability in effect sizes across these studies ($Q = 39.40$, $p = .0001$). With the exception of three studies, the effect size in the primary studies was significant (see the forest graph in figure 15).

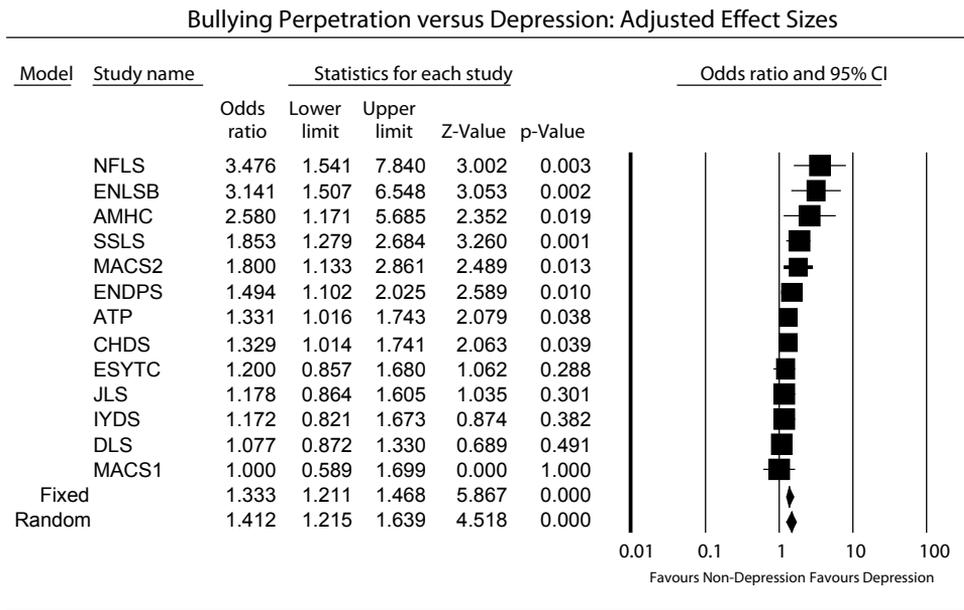
When controlling for covariates, the adjusted summary effect size was reduced to $OR = 1.41$, but this was still highly significant (95% CI: 1.22 – 1.64, $z = 4.52$). This OR indicates a moderate relationship between bullying perpetration and later depression. For example, if a quarter of children were bullies and a quarter were depressed, this value of the OR would correspond to 30.0% of bullies becoming depressed, compared with 23.3% of non-bullies. Thus, being a bully increased the risk of being depressed (even after controlling for other childhood risk factors) by about 30%. Figure 16 shows the forest graph for adjusted effect sizes. While all these effect sizes were in the expected direction, five were not statistically significant.

Figure 15. Unadjusted Effect Sizes for Bullying Perpetration versus Depression



Meta Analysis of Longitudinal Studies

Figure 16. Adjusted Effect Sizes for Bullying Perpetration versus Depression



Meta Analysis of Longitudinal Studies

6.2 Moderator Analyses

For the adjusted summary effect size, various moderators were investigated to explain the significant heterogeneity in effect sizes across studies ($Q = 24.98, p = .015$). These included the number of covariates controlled for at baseline (range: 1 – 20; $M = 6.62$; $SD = 5.62$), the age at which school bullying was measured (range: 8.00 – 15.54; $M = 11.76$; $SD = 2.73$), the age of participants when outcome measures were taken (range: 10.00 – 32.00; $M = 17.93$; $SD = 6.19$) and the length of the follow-up period, measured in years (range: 0.42 – 24.00; $M = 6.17$; $SD = 6.67$).

The age at which bullying was measured was positively associated with the effect size, but the regression coefficient was not statistically significant ($B = .032, SE = .020, p = .117$). The length of the follow-up period was negatively associated with the effect size and was close to significance ($B = -.009, SE = .006, p = .082$). The age of the study participants when outcome measures were taken was negatively related to the effect size but was not significant ($B = -.010, SE = .007, p = .127$). The above two negative relationships suggest a tendency for bullying perpetration to be more strongly related to depression in the short-term. The relationship between the number of covariates controlled for and the effect size was in the expected negative direction but not significant ($B = -.006, SE = .008, p = .435$).

As table 7 shows, all 13 studies with unadjusted and adjusted effect sizes for bullying perpetration versus depression were based on a longitudinal prospective design. Therefore, it was not possible to carry out a moderator analysis to investigate research design.

6.3 Publication Bias Analyses

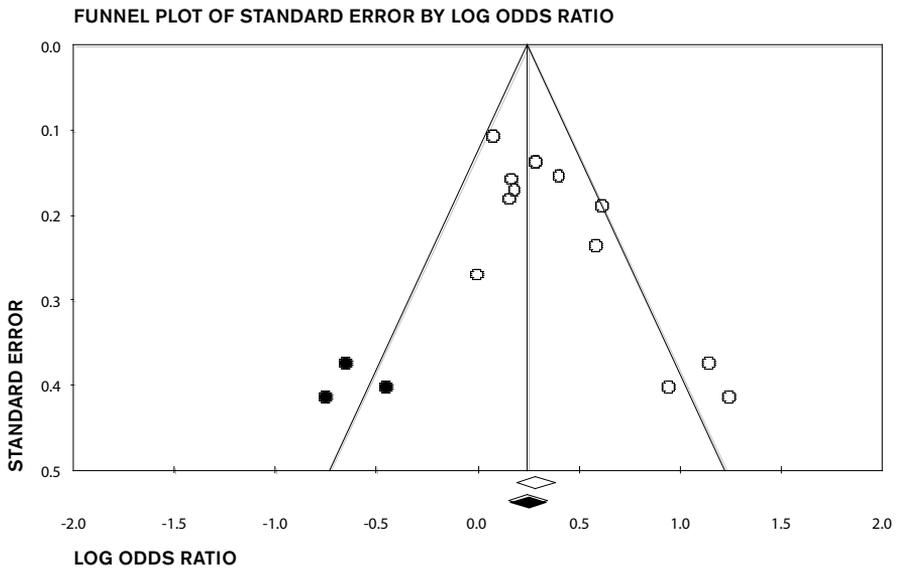
Figure 17 shows that the Duval and Tweedie's Trim-and-Fill procedure slightly shifted the imputed summary effect size (represented by a solid black diamond). Under the fixed effects model the point estimate and 95% confidence interval for the combined studies was 1.33318 (95% CI: 1.21106, 1.46761). Using Trim and Fill the imputed point estimate was 1.28142 (95% CI: 1.16653, 1.40763). Under the random effects model the point estimate and 95% confidence interval for the combined studies was 1.41154 (95% CI: 1.21548, 1.63922). Using Trim and Fill the imputed point estimate is 1.30555 (95% CI: 1.10147, 1.54743). These findings may suggest a slight overestimation of the true effect in our meta-analysis.

However, Rosenthal's Fail-Safe N test suggests differently. Here, the fail-safe N was 144. This means that we would need to locate and include 144 'null' studies in order for the combined 2-tailed p-value to exceed 0.050. Put another way, 11 missing studies for every observed study would be required for the effect to be nullified. It

is impossible that we have missed so many studies, or that so many were carried out but not published.

Finally, we conducted the Begg and Mazumdar rank correlation test. In this case Kendall's tau b (corrected for ties, if any) is 0.42308, with a 1-tailed p-value (recommended) of 0.02204 or a 2-tailed p-value of 0.04408 (based on continuity-corrected normal approximation). Overall, in spite of the slight tendency in the Trim and Fill procedure, the findings suggest that there is no publication bias in our findings.

Figure 17. Funnel Plot of Standard Error by Log Odds Ratio with Actual and Imputed Summary Effect Size (Bullying Perpetration versus Depression)



7. Conclusions and Recommendations

The results of our systematic review and meta-analyses suggest that there are long-term detrimental effects of bullying perpetration and victimization on later offending and depression. This was even true when confounded variables that are risks for bullying and victimization and for the undesirable outcomes were controlled for. Therefore, it can be concluded that bullying perpetration and victimization have independent effects on later psychosocial development. This is the first time that this conclusion is not based on a few selected primary studies and narrative reviews, but on comprehensive meta-analyses that include new data from a substantial body of yet unpublished research. Although the findings of the primary studies were not homogeneous, our summary effect sizes showed relatively narrow confidence intervals. Our findings also remained robust after we carried out sensitivity analyses testing potential publication bias.

The strongest effect was found for bullying perpetration versus later offending, but bullying victimization was also substantially related to later depression. Bullying increased the risk of later offending by more than half, and being bullied increased the risk of later depression by about half. The relationship between bullying perpetration and later offending may reflect an underlying disposition for antisocial behaviour that has different manifestations over time (Farrington, 1993; Lösel & Bliesener, 2003; Olweus, 1993). However, as the relation remained after controlling for other childhood risk factors, bullying seems to be a unique risk marker or may even perhaps be followed by an increased risk of offending (over and about the persistence of an underlying antisocial tendency). The underlying mechanisms may arise from the reinforcement obtained by dominating others and from the development of an identity as a 'bully' that persists beyond the school context. Similarly, the substantial adjusted effect size for victimization versus later de-

pression suggests that the frequent internalizing symptoms of victims may not only be a trigger for being bullied, but a psychological consequence. Victimization may lead to increased depression after controlling for the persistence of underlying tendencies.

Although the above-mentioned two meta-analyses revealed the strongest predictions, it is important to note that bullying perpetration was also significantly related to later depression and bullying victimization was nearly significantly related to later offending (when other childhood risk factors were controlled for). Bullying increased the risk of later depression by 30%, but victimization only increased the risk of later offending by 10%. The relationship between bullying perpetration and later depression may reflect the comorbidity of mental health problems and risk and protective mechanisms in development (Lösel & Farrington, 2011). Although internalizing problems can have a protective effect on the onset of offending (Loeber et al., 2008; Schwartz et al., 1996), they seem to be a risk factor for later onset (Zara & Farrington, 2009) and may hinder desistance in already antisocial youngsters (Loeber et al., 2008). Possibly, bullying leads to life failure and life failure leads to depression. The weak relationship between victimization and later offending may be driven by those children who show high levels of both victimization and bullying perpetration. These bully/victims are less clearly defined and frequent than 'pure' bullies and victims and often exhibit more reactive aggression (Lösel & Bliesener, 2003; Olweus, 1993). However, they seem to have similar problems of self-control as typical bullies.

Although the main findings of our meta-analyses were robust, there was always significant heterogeneity across primary studies. This could partly be explained in our moderator analyses. For example, there was some tendency for there to be smaller adjusted effect sizes when more childhood risk factors were controlled for. Effect sizes were a little smaller when outcomes were measured at older ages (or after longer follow up periods). This suggests that the effects of bullying perpetration and victimization may reduce over time.

The effect size variation may also be partly due to the method of measurement of bullying/victimization and offending/depression. For example, although we had clearly defined our inclusion criteria, 'offending' and 'depression' were measured differently across studies. As only a small number of studies included data from different informants (e.g. Averdijk et al., 2011; Farrington et al., 2011), we could not analyse such moderator effects in a systematic manner. The same applies to other measurement issues. For example, offending could have been defined as juvenile delinquency or serious criminality. A similar range of definitions existed for depression; the measurements often addressed depressive symptoms but were not usually clinical diagnoses. As mentioned in the introduction,

our review focussed on the overall school population (i.e. on children from the community) since we did not include studies based on clinic samples or studies in which participants were incarcerated or institutionalized youth.

Differences in sample characteristics were also relevant in explaining the variation of effect sizes. For example, in the Bender and Lösel (2011) study, the findings stem from a relatively small subsample that contained an over-sampling of typical bullies and victims. Such a large proportion of extreme cases would be expected to lead to larger effect sizes than in studies with population-based representative samples (such as the Nationwide Finnish 1981 Cohort Study). The gender composition of the samples may have also played a role in the effect size variation. However, not all studies included boys and girls and very few studies presented gender-specific results (e.g. Lösel & Bender, 2011). The ethnic-specific effect of school bullying on later adverse outcomes is another topic that could not be explored. To investigate and disentangle the impact of these and other variables, we need more longitudinal studies with a sound control for childhood risk factors. However, one should note that the lack of a sufficient number of studies with various different characteristics is a typical problem in meta-analyses (Lipsey, 2003).

In spite of such gaps in knowledge, our overall robust findings have clear implications for practice, policy making and research. For example, they provide sound information for teachers, school psychologists, social workers, intervention planners, mental health experts, and other professionals who are confronted with school bullying and its consequences. The substantial links between school bullying/victimization and later offending/depression give a strong voice to anti-bullying agencies and (re-)establish the moral imperative for school communities to create a school ethos that is as much free of bullying as possible. This could be achieved by effective anti-bullying policies in schools that are legally binding (as opposed to the current situation in most European school systems; see Ananiadou & Smith, 2002) and are followed with scrutiny regarding the real content of such policies and their implementation (Farrington & Ttofi, 2009).

Parents need also to become fully aware of the serious and long-term negative impacts of school bullying. Bullying should not be perceived as ‘one of those childhood experiences for the grownup world’, because this would hinder child protection and indirectly perpetuate the problem. Previous research (Joffre-Velazquez et al., 2011) has found that children whose parents consider bullying to be a normal problem are almost six times more likely to be victimized than children whose parents take bullying incidents more seriously. We are hopeful that the findings of the present report will help these parents and teachers become more aware of the negative effect of school bullying.

The consistent and substantial relations between school bullying and later outcomes are also relevant from an economic perspective. Research has shown that a single case of serious and long-term youth crime could cost society up to five million US Dollars (Cohen & Piquero, 2009). Thus, interventions and social policies that effectively reduce bullying could be viewed as an early form of crime prevention that help to interrupt costly long-term antisocial developments. Similarly, measures that reduce bullying victimization could save later costs for mental health care. We have shown elsewhere what types of anti-bullying programmes are most effective in reducing bullying perpetration and victimization (Farrington & Ttofi 2009; Ttofi and Farrington, 2011). However, the majority of the intervention studies had only relatively short follow-up periods. The need for more long-term evaluations is suggested by the findings of the present report.

Finally, our study may also be of interest with regards to its methodology. To date, systematic reviews in criminology have focused primarily on measures of prevention and intervention. In this case, randomized experiments and sound quasi-experimental designs provide the best evidence on causal effects. However, there are also many naturally occurring (i.e. non-manipulated) origins of delinquency or other behavioural problems. Typical examples are growing up in a broken home, brain damage, parental incarceration or, with reference to the current report, bullying victimization (Murray et al., 2009). Such events cannot be randomly assigned and manipulated because of ethical and other reasons (Petrosino, 2003), but may be highly relevant for pathways into or out of crime. Systematic reviews of longitudinal studies which control for confounded variables can give some hints on whether variables are simple correlational risk factors, risk markers or causal risk factors (Kraemer et al., 2005). More research is needed to investigate the effects of bullying perpetration and victimization after controlling for the persistence of underlying tendencies. Nevertheless, our research suggests that bullying perpetration may lead to an increase in later offending, and that bullying victimization may lead to an increase in later depression.

References⁶⁸

- Ananiadou, K. & Smith, P.K. (2002). Legal requirements and nationally circulated materials against school bullying in European countries. *Criminal Justice*, 2 (4), 471–491.
- * Arseneault, L. (2011). *Adjusted effect sizes for the E-Risk Longitudinal Study*. Email communication, January 21, 2011.
- Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and mental health problems: ‘Much ado about nothing’? *Psychological Medicine*, 40, 717–729.
- * Averdijk, M., Muller, B., Eisner, M., & Ribeaud, D. (2011). Bullying victimization and later anxiety and depression among pre-adolescents in Switzerland. *Journal of Aggression, Conflict and Peace Research*, 3, 103–109.
- * Azzuzi, S. & Killias, M. (2010). *Long-term outcomes of active and passive bullying during childhood: Suicide attempts, victimization and adult violence. A research note*. Unpublished manuscript received via email communication with Professor Killias, June 11, 2010.
- * Barbarin, O.A. (1999). Social risks and psychological adjustment: A comparison of African American and South African children. *Child Development*, 70, 1348–1359.
- * Barker, E.D., Arseneault, L., Brendgen, M., Fontaine, N., & Maughan, B. (2008). Joint development of bullying and victimization in adolescence: Relations to delinquency and self-harm. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47 (9), 1030–1038.
- * Bender, D. & Lösel, F. (2011). Bullying at school as a predictor of delinquency, violence and other antisocial behavior in adulthood. *Criminal Behaviour and Mental Health*, 21, 99–106.
- * Berry, K. & Hunt, C. (2009). Evaluation of an intervention program for anxious adolescent boys who are bullied at school. *Journal of Adolescent Health*, 45, 376–382.

⁶⁸ References with an asterisk indicate reports included in the systematic reviews but not necessarily in the meta-analyses.

- * Bijleveld, C., van der Geest, V., & Hendriks, J. (2011). Bullying and (re)offending: Results from three samples in the Netherlands. *Criminal Behaviour and Mental Health*, 21, 14–150.
- Blumstein, A., Cohen, J., Roth, J.A., & Visser, C.A. (1986). *Criminal careers and 'career criminals'*. Washington DC: National Academy Press.
- * Bond, L., Carlin, J.B., Thomas, L., Rubin, K., & Patton, G. (2001). Does bullying cause emotional problems? A prospective study of young teenagers. *British Medical Journal*, 323, 480–84.
- * Boulton, M.J., Smith, P.K., & Cowie, H. (2010). Short-term longitudinal relationships between children's peer victimization/bullying experiences and self-perceptions: Evidence for reciprocity. *School Psychology International*, 31 (3), 296–311.
- * Bowes, L., Arseneault, L., Maughan, B., Taylor, A., Caspi, A., Moffitt, T.E. (2009). School neighborhood, and family factors are associated with children's bullying involvement: A nationally representative longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48 (5), 545–553.
- Bowes, L., Maughan, B., Caspi, A., Moffitt, T.E., & Arseneault, L. (2010). Families promote emotional and behavioural resilience to bullying: Evidence of an environmental effect. *Journal of Child Psychology and Psychiatry*, 51 (7), 809–817.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- * Carlisle, N. & Rofes, E. (2007). School bullying: Do adult survivors perceive long-term effects? *Traumatology*, 13(1), 16–26.
- Cohen M.A. & Piquero A.R. (2009). New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology*, 25, 25–49.
- * DeRosier, M.E. (2004). Building relationships and combating bullying: Effectiveness of a school-based social skills group intervention. *Journal of Clinical Child and Adolescent Psychology*, 33 (1), 196–201.
- * DeRosier, M.E. (2007). Peer-rejected and bullied children: A safe schools initiative for elementary school students. In J.E. Zins, M.J. Elias, & C.A. Maher (Eds.), *Bullying, victimization and peer harassment: A handbook of prevention and intervention* (pp. 257–276). New York: The Haworth Press.
- * DeRosier, M.E. & Marcus, S.R. (2005). Building relationships and combating bullying: Effectiveness of S.S.GRIN at one-year-follow-up. *Journal of Clinical Child and Adolescent Psychology*, 34 (1), 140–150.
- * Due, P., Damsgaard, M.T., Lund, R., & Holtstein, B.E. (2009). Is bullying equally harmful for rich and poor children? A study of bullying and depression from age 15 to 27. *European Journal of Public Health*, 19, 464–469.

- * Espelage, D.L., Bosworth, K., & Simon, T.R. (2001). Short-term stability and prospective correlates of bullying in middle-school students: An examination of potential demographic, psychosocial, and environmental influences. *Violence and Victims*, 16, 411–426.
- Farrington, D.P. (1993). Understanding and preventing bullying. In M. Tonry (Ed.), *Crime and Justice*, Vol. 17 (pp. 381 – 458). Chicago: University of Chicago Press.
- Farrington, D.P. & Ttofi, M.M. (2009). *School-based programs to reduce bullying and victimization*. Campbell Systematic Reviews, 2009:6.
- * Farrington, D.P. & Ttofi, M.M. (2011). Bullying as a predictor of offending, violence and later life outcomes. *Criminal Behaviour and Mental Health*, 21, 90–98.
- * Farrington, D.P., Loeber, R., Stallings, R., & Ttofi, M.M. (2011a). Bullying perpetration and victimization as predictors of delinquency and depression in the Pittsburgh Youth Study. *Journal of Aggression, Conflict and Peace Research*, 3, 74–81.
- Farrington, D.P., Ttofi, M.M., & Lösel, F. (2011b). Editorial: School bullying and later offending. *Criminal Behaviour and Mental Health*, 21, 77– 9.
- * Fekkes, M. Pijpers, F. I. M. Fredriks, A. M. Vogels, T., & Verloove-Vanhorick, S. P. (2006). Do bullied children get ill, or do ill children get bullied? A prospective cohort study on the relationship between bullying and health-related symptoms. *Pediatrics*, 117, 1568–1574.
- * Ferguson, C.J. (2011). Video games and youth violence: A prospective analysis in adolescents. *Journal of Youth and Adolescence*, 40, 377–391.
- * Gibb, S.J., Horwood, L.J., & Fergusson, D.M. (2011). Bullying victimization/perpetration in childhood and later adjustment: Findings from a 30 year longitudinal study. *Journal of Aggression, Conflict and Peace Research*, 3 (2), 82– 8.
- Gilmartin, B.G. (1987). Peer group antecedents of severe love-shyness in males. *Journal of Personality*, 55, 467–489.
- * Gladstone, G.L. & Parker, G.B. (2006). Is behavioral inhibition a risk factor for depression? *Journal of Affective Disorders*, 95, 85–94.
- * Gladstone, G.L., Parker, G.B., & Malhi, G.S. (2006). Do bullied children become anxious and depressed adults? A cross-sectional investigation of the correlates of bullying and anxious depression. *Journal of Nervous and Mental Disease*, 194, 201–208.
- * Grills, A. (2003). *Long-term relations among peer victimization and internalizing symptoms in children*. Phd Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

- * Haapasalo, J., Tremblay, R.E., Boulerice, B., & Vitaro, F. (2000). Relative advantages of person- and variable-based approaches for predicting problem behaviors from kindergarten assessments. *Journal of Quantitative Criminology*, 16 (2), 145–168.
- * Haas, H. (2001). *Aggression et victimizations*. Aaru: Sauerlander.
- * Haavisto, A., Sourander, A., Multimaki, P., Parkkola, K., Santalahti, P., Helenius, H., Nikolakaros, G., Kumpulainen, K., Moilanen, I., Piha, J., Aronen, E., Puura, K., Linna, S-L., & Almqvist, F. (2004). Factors associated with depressive symptoms among 18-year-old boys: A prospective 10-year follow-up study. *Journal of Affective Disorders*, 83, 143– 54.
- * Hemphill, S.A., Kotevski, A., Herrenkohl, T.I., Bond, L., Kim, M.J., Toumbourou, J.W., & Catalano, R.F. (2011). Longitudinal consequences of adolescent bullying perpetration and victimization: A study of students in Victoria, Australia. *Criminal Behaviour and Mental Health*, 21, 107–116.
- * Henry, D., Guerra, N., Tolan, PL, Huesmann, L.R., VanAcker, R., & Eron, L. (2010). *Metropolitan Area Child Study, 1991–1998: Unpublished Raw Data for Cohort 1*. Email communication with Professor Henry, March 24, 2010.
- * Henry, D., Guerra, N., Tolan, PL, Huesmann, L.R., VanAcker, R., & Eron, L. (2010). *Metropolitan Area Child Study, 1991–1998: Unpublished Raw Data for Cohort 2*. Email communication with Professor Henry, March 24, 2010.
- Herrenkohl T.I., McMorris B.J., Catalano, R.F., Abbott, R.D., Hemphill, S.A., & Toumbourou, J.W. (2007). Risk factors for violence and relational aggression in adolescence. *Journal of Interpersonal Violence*, 22, 386–405.
- * Hidaka, Y. & Operario, D. (2006). Attempted suicide, psychological health and exposure to harassment among Japanese homosexual, bisexual or other men questioning their sexual orientation recruited via the internet. *Journal of Epidemiology and Community Health*, 60, 962–967.
- * Holmes, S.R. & Brandenburg-Ayres, S.J. (1998). Bullying behavior in school: A predictor of later gang involvement. *Journal of Gang Research*, 5(2), 1–6.
- * Jiang, D., Walsh, M., & Augimeri, L.K. (2011). The linkage between childhood bullying behaviour and future offending. *Criminal Behaviour and Mental Health*, 21, 128–135.
- Joffre-Velazquez, V.M., Garcia-Maldonado, G., Saldivar-Gonzalez, A.H., Martinez-Perales, G., Lin-Ochoa, D., Quintanar-Martinez, S., & Villasana-Guerra, A. (2011). Bullying in junior high school students: General characteristics and associated risk factors. *Boletin Medico del Hospital Infantil de Mexico*, 68, 177–185.

- * Jordanova, V., Stewart, R., Goldberg, D., Bebbington, P.E., Brugha, T., Singleton, N., Lindesay, J.E.B., Jenkins, R., Prince, M., & Meltzer, H. (2007). Age variation in life events and their relationship with common mental disorders in a national survey population. *Social Psychiatry and Psychiatric Epidemiology*, 42, 611–616.
- * Kaltiala-Heino, R., Frojd, S., & Martunen, M. (2010). Involvement in bullying and depression in a 2-year follow-up in middle adolescence. *European Child and Adolescent Psychiatry*, 19 (1), 45–55.
- * Kendrick, K. & Hagan, S. (2010). *Unpublished data analyses for the Seven Schools Longitudinal Study*. Email communication with Kristin Kenrick, February 22 and 26, 2010.
- * Kim, M.J., Catalano, R.F., Haggerty, K.P., & Abbott, R.D. (2011). Bullying at elementary school and problem behaviour in young adulthood: A study of bullying, violence and substance use from age 11 to age 21. *Criminal Behaviour and Mental Health*, 21, 136–144.
- * Kimber, B., Sandell, R., & Bremberg, S. (2008a). Social and emotional training in Swedish classrooms for the promotion of mental health: Results from an effectiveness study in Sweden. *Health Promotion International*, 23, 134–143.
- * Kimber, B., Sandell, R., & Bremberg, S. (2008b). Social and emotional training in Swedish schools for the promotion of mental health: An effectiveness study of 5 years of intervention. *Health Education Research*, 23, 931–940.
- * Klomek, A.B., Sourander, A., Kumpulainen, K., Piha, J., Tamminen, T., Moilanen, I., Almqvist, F., & Gould, M.S. (2008). Childhood bullying as a risk for later depression and suicidal ideation among Finnish males. *Journal of Affective Disorders*, 109, 47–55.
- * Klomek, A.B., Sourander, A., Niemela, S., Kumpulainen, K., Piha, J., Tanninen, T., Almqvist, F., & Gould, M.S. (2009). Childhood bullying behaviors as a risk for suicide attempts and completed suicides: A population-based cohort study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 254–261.
- Kraemer, H.C., Lowe, K.K., & Kupfer, D.J. (2005). *To your health: How to understand what research tells us about risks*. New York: Oxford University Press.
- * Kumpulainen, K. & Rasanen, E. (2000). Children involved in bullying at elementary school age: Their psychiatric symptoms and deviance in adolescence. *Child Abuse and Neglect*, 24 (12), 1567–1577.
- * Kumpulainen, K. & Roine, S. (2002). Depressive symptoms at the age of 12 years and future heavy alcohol use. *Addictive Behaviors*, 27, 425–436.

- * Kumpulainen, K., Rasanen, E., Henttonen, I., Hamalainen, M., & Roine, S. (2000). The persistence of psychiatric deviance from the age of 8 to the age of 15 years. *Social Psychiatry and Psychiatric Epidemiology*, 35, 5–11.
- * Kumpulainen, K., Rasanen, E., & Puura, K. (2001). Psychiatric disorders and the use of mental health services among children involved in bullying. *Aggressive Behavior*, 27, 102–110.
- Lipsey, M.W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science*, 587, 69–81.
- Loeber, R., Farrington, D.P., Stouthamer-Loeber, M., & White, H.R. (2008). *Violence and serious theft: Development and prediction from childhood to adulthood*. New York: Routledge.
- * Lösel, F. & Bender, D. (2011). Emotional and antisocial outcomes of bullying and victimization at school: A follow-up study from childhood to adolescence. *Journal of Aggression, Conflict and Peace Research*, 3, 89–96.
- * Lösel, F. & Bliesener, T. (2003). *Aggression und delinquenz unter jugendlichen: Untersuchungen von kognitiven und sozialen Bedingungen*. Hermann: Luchterhand.
- Lösel, F. & Farrington, D.P. (2011). *Promotive and protective factors in the development of youth violence*. Report prepared for the U.S. Centers of Disease Control and Prevention.
- * Lösel, F., Bender, D., Fehn, J., & Schulze, V. (2008). *Long-term follow up of school bullies: Social and personality development from age 15 to 25 years*. Paper presented at the 8th Annual Conference of the European Society of Criminology, Edinburgh (September 4).
- * Lund, R., Nielsen, K.K., Hansen, D.H., Kriegbaum, M., Molbo, D., Due, P., & Chistensen, U. (2008). Exposure to bullying at school and depression in adulthood: A study of Danish men born in 1953. *The European Journal of Public Health*, 19 (1), 111–116.
- * Luukkonen, A.H., Riala, K., Hakko, H., & Rasanen, P. (2011). Bullying behaviour and criminality: A population-based follow-up study of adolescent psychiatric inpatients in Northern Finland. *Forensic Science International*, 207, 106–110.
- * Marsh, H.W., Parada, R.H., Craven, R.G., & Finger, L. (2004). In the looking glass: A reciprocal effects model elucidating the complex nature of bullying, psychological determinants, and the central role of self-concept. In C.E. Sanders & G.D. Phye (Eds.), *Bullying: Implications for the classroom* (pp. 63–109). London: Elsevier.
- * Matsui, T., Tsuzuki, Y., Kakuyama, T. & Onglatco, M. L. (1996). Long term outcomes of early victimization by peers among Japanese male university students: Model of a vicious cycle. *Psychological Reports*, 79, 711–720.

- * McGee, T.R., Scott, J.G., McGrath, J.J., Williams, G.M., O'Callaghan, M., Bor, W., & Najman, J.M. (2011). Youth adult problem behavior outcomes of adolescent bullying. *Journal of Aggression, Conflict and Peace Research*, 3, 110–114.
- * McVie, S. (2010). *Data analyses for the Edinburgh Study of Youth Transitions and Crime: Unpublished data analyses*. Email communication with Susan McVie, December 22, 2010.
- * Moffitt, T.E., Houts, R., & Harrington, H. (2010). *Unpublished data based on the Dunedin Longitudinal Study*. Email communication with Renate Houts, July 22, 2010.
- * Moon, B., Hwang, H-W., & McCluskey, J.D. (2011). Causes of school bullying: Empirical test of a general theory of crime, differential association, and general strain theory. *Crime and Delinquency*, 57, 849–877.
- Murray J., Farrington D.P., Eisner M.P. (2009). Drawing conclusions about causes from systematic reviews of risk factors: The Cambridge Quality Checklists. *Journal of Experimental Criminology*, 5, 1–23.
- * Nishino, Y., Ujiie, T., Nonomiya, K., Igarashi, A., Inoue, H., & Yamamoto, C. (2009). A longitudinal study of the trajectories of deviant behavior among junior high school students. *Shinrigaku Kenkyu*, 80 (1), 17–24.
- * Nishino, Y., Ujiie, T., Nonomiya, K., Igarashi, A., Inoue, H., & Yamamoto, C. (2011). Longitudinal study of the relationship between victimization and later emotional problems among Japanese junior high school students. *Journal of Aggression, Conflict and Peace Research*, 3, 115–121.
- * Nrugham, L. (2010). *Associates and predictors of attempted suicide among depressed adolescents: A 6-year prospective study*. PhD Thesis, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway.
- * Nrugham, L., Larsson, B., Sund, A.M. (2008). Predictors of suicidal acts across adolescence: Influences of familial, peer and individual factors. *Journal of Affective Disorders*, 109, 35– 5.
- * Olweus, D. (1991). Bully/victim problems among school children: Basic facts and effects of a school-based intervention program. In D.J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 411–448). Hillsdale, NJ: Erlbaum.
- * Olweus, D. (1993a). *Bullying at school: what we know and what we can do*. Oxford, England: Blackwell.
- * Olweus, D. (1993b). Bully/victim problems among schoolchildren: long-term consequences and an effective intervention program. In S. Hodgins (Ed.), *Mental disorder and crime* (pp. 317–349). Thousand Oaks, CA, US: Sage Publications.

- * Olweus, D. (1993c). Victimization by peers: Antecedents and long-term outcomes. In K.H. Rubin & J.B. Asendorff (Eds.), *Social withdrawal, inhibition and shyness* (pp. 315–342). Hillsdale, NJ: Lawrence Erlbaum.
- * Olweus, D. (1994a). Annotation: Bullying at school: basic facts and effects of a school based intervention program. *Journal of Child Psychology and Psychiatry*, 35, 1171–1190.
- * Olweus, D. (1994b). Bullying at school: long-term outcome for the victims and an effective school-based intervention program. In L.R. Huesmann (Ed.), *Aggressive Behavior: current perspectives* (pp. 97–130). New York: Plenum.
- * Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education*, 12, 495–510.
- * Olweus, D. (2011). Bullying at school and later criminality: Findings from three Swedish community samples of males. *Criminal Behaviour and Mental Health*, 21, 151–156.
- O'Moore, M., Seigne, E., McGuire, L., & Smith, M. (1998). Victims of workplace bullying in Ireland. *The Irish Journal of Psychology*, 15, 2–5, 345–357.
- * Ozdemir, M. & Stattin, H. (2011). Bullies, victims and bully-victims: A longitudinal examination of the effects of bullying-victimization experiences on youth well-being. *Journal of Aggression, Conflict and Peace Research*, 3, 97–102.
- * Parada, R.H. (2006). *School bullying: psychosocial determinants and effective intervention*. Unpublished PhD thesis, University of Western Sydney.
- * Parada, R.H., Craven, R.G., & Marsch, H.W. (2008). The Beyond Bullying secondary school program. In H.W. Marsh, R.G. Craven, & D.M. McInerney (Eds.), *Self-processes, learning, and enabling human potential: Dynamic new approaches* (pp. 373–426). Charlotte, NC: Information Age Publishing.
- * Patton, G.C., Olsson, C., Bond, L., Toumbourou, J.W., Carlin, J.B., Hemphill, S.A., & Catalano, R.F. (2008). Predicting female depression across puberty: A two-nation longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 1424–432.
- Petrosino, A. (2003). Estimates of randomized controlled trials across six areas of childhood intervention: a bibliometric analysis. *Annals of the American Academy of Political and Social Sciences*, 589, 190–202.
- * Pirkola, S., Isometsa, E., Aro, H., Kestila, L., Hamalainen, J., Veijola, J., Kiviruusu, O., & Lonnqvist, J. (2005). Childhood adversities as risk factors for adult mental disorders: Results from the Health 2000 Study. *Social Psychiatry and Psychiatric Epidemiology*, 40, 769–777.

- * Pulkkinen, L. & Tremblay, R.E. (1992). Patterns of boys' social adjustment in two cultures and at different ages: A longitudinal perspective. *International Journal of Behavioral Development*, 15 (4), 527–553.
- * Renda, J., Vassallo, S., & Edwards, B. (2011). Bullying in early adolescence and its association with anti-social behaviour, criminality and violence 6 and 10 years later. *Criminal Behaviour and Mental Health*, 21, 117–127.
- * Rigby, K. (1999). Peer victimization at school and the health of secondary school students. *British Journal of Educational Psychology*, 69(1), 95–104.
- * Rigby, K. (2001). Health consequences of bullying and its prevention in schools. In J. Juvonen & G. Sandra (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 310–331). New York: Guilford.
- * Rivers, I. (1999). *The psycho-social correlates and long-term implications of bullying at school for lesbians, gay men and bisexual men and women*. Unpublished PhD thesis, University of Surrey.
- * Rivers, I. (2001). The bullying of sexual minorities at school: Its nature and long term correlates. *Educational and Child Psychology*, 18 (1), 32–46.
- * Rivers, I. (2004). Recollections of bullying at school and their long-term implications for lesbians, gay men and bisexuals. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 25, 169–175.
- * Rivers, I. & Cowie, H.A. (2006). Bullying and homophobia in UK schools: A perspective on factors affecting resilience and recovery. *Journal of Gay and Lesbian Issues in Education*, 3, 11–43.
- * Roeger, L., Allison, S., Korossy-Horwood, R., Eckert, K.A., & Goldney, R.D. (2010). Is a history of school bullying victimization associated with adult suicidal ideation? A south Australian population-based observational study. *Journal of Nervous and Mental Disease*, 198, 728–733.
- * Roth, D., Coles, M. & Heimberg, R. (2002). The relationship between memories for childhood teasing and anxiety and depression in adulthood. *Anxiety Disorders*, 16, 149–164.
- * Salmivalli, C. (2010). KiVa AntiBullying Programme: Follow-up Intervention Study: Unpublished Data. Email communication with Christina Salmivalli, March 29, 2010.
- Schwartz, C.E., Snidman, N., & Kagan, J. (1996). Early childhood temperament as a determinant of externalizing behavior in adolescence. *Development and Psychopathology*, 8, 527–537.
- * Shakoor, S., Jaffee, S.R., Andreou, P., Bowes, L., Ambler, A.P., Caspi, A., Moffit, T.E., & Arsenuault, L. (2011). Mothers and children as informants of bullying victimization: Results from an epidemiological cohort of children. *Journal of Abnormal Child Psychology*, 39, 379–387.

- * Shelley, S.D. (2009). *Thinking and doing: Attributions and coping of children and their friends that are associated with the continuity of victimization and bullying*. Unpublished PhD thesis, Department of Psychology, Queen's University, Kingston, Ontario: Canada.
- * Shelley, S.D. & Craig, W. (2010). Attributions and coping styles in reducing victimization. *Canadian Journal of School Psychology*, 25 (1), 84–100.
- * Singer, M. (2002). *Long-term effects of school bullying: Consequences of being bullied and the influence of sense of humour and spirituality*. Unpublished PhD thesis, Goldsmiths College, University of London.
- * Smith, D.J. & Ecob, R. (2007). An investigation into causal links between victimization and offending in adolescents. *The British Journal of Sociology*, 58 (4), 633–659.
- Smith, P.K. (1997). Bullying in life-span perspective: What can studies of school bullying and workplace bullying learn from each other? *Journal of Community and Applied Social Psychology*, 7, 249–255.
- Smith, P.K. & Brain, P. (2000). Bullying in schools: Lessons from two decades of research. *Aggressive Behavior*, 26, 1–9.
- * Smith, P.K., Talamelli, L., Cowie, H., Naylor, P., Chauhan, P. (2004). Profiles of non-victims, escaped victims, continuing victims and new victims of school bullying. *British Journal of Educational Psychology*, 74(4), 565–581.
- * Sourander, A., Elonheimo, H., Niemela, S., Nuutila, A., Helenius, H., Sillanmaki, L., Piha, J., Tamminen, T., Kumpulainen, K., Moilenen, I., & Almqvist, F. (2006). Childhood predictors of male criminality: A prospective population-based follow-up study from age 8 to late adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 578–586.
- * Sourander, A., Helstela, L., Helenius, H., & Piha, J. (2000). Persistence of bullying from childhood to adolescence: A longitudinal 8-year follow-up study. *Child Abuse and Neglect*, 24, 873–881.
- * Sourander, A., Jensen, P., Ronning, J. A., Elonheimo, H., Niemela, S., Helenius, H., Kumpulainen, K., Piha, J., Tamminen, T., Moilanen, I., & Almqvist, F. (2007a). Childhood bullies and victims and their risk of criminality in late adolescence: The Finnish from a boy to a man study. *Archives of Pediatrics and Adolescent Medicine*, 161(6), 546–552.
- * Sourander, A., Jensen, P., Ronning, J., Niemela, S., Helenius, H., Sillanmaki, L., Kumpulainen, K., Piha, J., Tamminen, T., Moilanen, I. & Almqvist, F. (2007b). What is the early adulthood outcome of boys who bully or are bullied in childhood? The Finnish “From a boy to a man” study. *Pediatrics*, 120, 397–404.

- * Sourander, A., Klomek, A.B., Kumpulainen, K., Puustjarvi, A., Elonheimo, H., Ristkari, T., Tamminen, T., Moilanen, I., Piha, J., & Ronning, J.A. (2011). Bullying at age eight and criminality in adulthood: Findings from the Finnish Nationwide 1981 Birth Cohort Study. *Social Psychiatry and Psychiatric Epidemiology*, 46, 1211–1219.
- * Sourander, A., Ronning, J., Brunstein-Klomek, A., Gyllenberg, D., Kumpulainen, K., Niemela, S., Helenius, H., Sillanmaki, L., Ristkari, T., Tamminen, T., Moilanen, I., Piha, J. & Almqvist, F. (2009). Childhood bullying behavior and later psychiatric hospital and psychopharmacologic treatment: Findings from the Finnish 1981 birth cohort study. *Archives of General Psychiatry*, 66, 1005–1012.
- * Staubli, S. & Killias, M. (2011). Long-term outcomes of passive bullying during childhood: Suicide attempts, victimization and offending. *European Journal of Criminology*, 8, 377–385.
- * Taylor, R.J. (2006). *An evaluation of Owing Up: Impact of perceptions of relational aggression, bullying and victimization*. Unpublished PhD thesis, New York, USA: Hofstra University.
- * Tremblay, R.E. & Haapasalo, J. (1998). Boys' behavioral patterns in kindergarten predict male delinquency, withdrawal, and school adjustment. In J. Boros, I. Munnich, & M. Szegedi (Eds.), *Psychology and criminal justice: International review of theory and practice* (pp. 197–211). Berlin, New York: Walter de Gruyter.
- Ttofi, M. M. & Farrington, D. P. (2008) Bullying: short-term and long-term effects, and the importance of Defiance Theory in explanation and prevention. *Victims and Offenders*, 3 (2), 289–312.
- Ttofi, M.M. & Farrington, D.P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, 7, 27–56.
- Ttofi, M. M., Farrington, D. P., & Baldry, A. C. (2008). *Effectiveness of programmes to reduce school bullying: A systematic review*. Stockholm: Swedish National Council for Crime Prevention.
- Ttofi, M.M., Farrington, D.P., & Lösel, F. (2011a). Editorial: Health consequences of school bullying. *Journal of Aggression, Conflict and Peace Research*, 3 (2), 60–62.
- Ttofi, M.M., Farrington, D.P., Lösel, F., & Loeber, R. (2011b). Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Journal of Aggression, Conflict and Peace Research*, 3, 63–73.
- Ttofi, M.M., Farrington, D.P., Lösel, F., & Loeber, R. (2011c). The predictive efficiency of school bullying versus later offending: A systematic/meta-analytic review of longitudinal studies. *Criminal Behaviour and Mental Health*, 21, 80–89.

- * Vaillancourt, T., Duku, E., Becker, S., Schmidt, L.A., Nicol, J., Muir, C., & MacMillan, H. (2011). Peer victimization, depressive symptoms, and high salivary cortisol predict poorer memory in children. *Brain and Cognition*, 77, 191–199.
 - * White, N.A. & Loeber, R. (2008). Bullying and special education as predictors of serious delinquency. *Journal of Research in Crime and Delinquency*, 45 (4), 380–397.
 - * Wong, J.S. (2009). *No bullies allowed: Understanding peer victimization, the impacts on delinquency and the effectiveness of prevention programmes*. Unpublished PhD thesis. Pardee Rand Graduate School, USA.
- Zara, G. & Farrington, D.P. (2009). Childhood and adolescent predictors of late onset criminal careers. *Journal of Youth and Adolescence*, 38, 287–300.

Appendices

Appendix 1: Combining Data on Offending Outcomes

For the Australian Temperament Project (Renda et al., 2011), results on bullying perpetration versus offending are based on police/court contact. Results on victimization versus offending are based on the unpublished data provided via email. We have combined the effect sizes for property damage and shoplifting. The results of the unpublished data do not include the same follow-up period as the published data: they are based on offending at age 23–24 only. For the Cambridge Study in Delinquent Development (Farrington & Ttofi, 2011), we present results on bullying perpetration versus convictions, based on official record data. Results on bullying victimization were not included in the dataset. Preference was given to the most recent report over older ones (i.e. Farrington, 1993).

For the Christchurch Health and Development Study (Gibb et al., 2011), results on bullying perpetration versus offending are based on two separate self-report measures of property offending and arrest/convictions. Separate effect sizes are presented for middle childhood and adolescence. For the same study, effect sizes for victimization versus offending are based on the same outcome measures, but the predictor is only victimization in adolescence (and the authors also control for fewer covariates). For the Edinburgh Study of Youth Transitions and Crime (McVie, 2010), results on bullying perpetration versus offending are based on two separate self-report items on property theft and property damage; the same applies for victimization. Mean age at baseline and follow-up period is the same for both predictors. Preference was given to the most recent report over older ones (i.e. Barker et al., 2008; Smith & Ecob, 2007), which also did not offer adjusted effect sizes.

For the Erlangen-Nuremberg Development and Prevention Study (Lösel & Bender, 2011), a combined self-reported and mother-reported measure of delinquency was used, with the same covariates

and follow-up periods (for both bullying and victimization) included in the analyses. The same covariates and mean age at baseline and follow-up period was used for the association of bullying (perpetration and victimization) versus self-reported offending for the Erlangen-Nuremberg Longitudinal Study of Bullying (Bender & Lösel, 2011). We should mention that an older report for this study (Lösel & Bliesener, 2003) presents similar findings but was written in German. The most recent report of the study (Bender & Lösel, 2011) makes the study findings more available to international researchers as it is written in English. The Lösel et al. (2008) report is a conference presentation and the written version (i.e. slides for the presentation with data findings) provides similar findings to the most recent 2011 report.

Two manuscripts with offending outcomes report results from the Boy to a Man Finnish Longitudinal Study (Sourander et al., 2006, 2007a) and another one was based on the Finnish Cohort Longitudinal Study (Sourander et al., 2011), which includes data on both males and females. In the Sourander et al. (2007a) paper, unadjusted effect sizes were shown separately for bullying perpetration versus violent, property, traffic, drunk driving and drug offences (table 2: 549). The paper also shows adjusted effect sizes after controlling for parental education (table 1: 549) for '1-2 crimes' and 'more than two crimes'. We felt that we could not use these adjusted effect sizes (and compare them with a summary effect size combined across the 5 different types of offences) since we could not possibly know which of the five types of offences were included in the '1 to 2 crimes' and '2 or more' categories. In the Sourander et al. (2006) paper, the reverse pattern occurred: unadjusted effect sizes were shown for bullying perpetration versus '1-2 offences', for '3-5 offences' and for 'five or more offences' whereas adjusted effect sizes were shown separately for bullying versus violent, property, traffic, drunk driving and drug offences (five different adjusted effect sizes; see table 3: 583). With regard to victimization, the Sourander et al. (2006) paper does not show any adjusted effect sizes for bullying victimization at all, but it shows unadjusted effect sizes for victims of school bullying versus 1 – 2 crimes, 3 – 5 crimes and for more than 5 crimes. The Sourander et al. (2007a) paper shows adjusted effect sizes for 1 – 2 crimes and for more than 2 crimes after controlling for parental education. In the 2006 paper, bullying (perpetration and victimization) was self-rated (Sourander et al., 2006: 580), while in the 2007 paper bullying (perpetration and victimization) was either self-, teacher- or parent-rated (Sourander et al., 2007a: 547).

The latest paper by Sourander et al. (2011) shows both unadjusted and adjusted effect sizes for bullying (perpetration and victimization) versus offending. For the meta-analyses on bullying (perpetration/victimization) versus offending, we have chosen this more

recent paper over the other two because the data presented in the 2011 paper were based on a longer follow-up period and included results on both males and females. Results were shown separately for males and females and we have combined the relevant effect sizes. Because of small numbers, results for females are based on ‘sometimes’ or ‘frequent’ bullying (while for males the two categories are separated) and for ‘more than one crime’ as an outcome measure (while for the males an outcome measure of ‘more than 5 crimes’ is presented). Despite these differences, this 2011 paper is based on the most up-to-date results for the study and includes the female population as well (it is a population-based study). The relevant tables (Sourander et al., 2011: 1215 – 1216; see their tables 1 and 2) show unadjusted effect sizes but also provide the percentages within each category, based on which we have calculated the relevant unadjusted effect sizes. We have combined the ‘sometimes’ and ‘frequent’ bullying/victimization categories for males to make them comparable to the relevant effect sizes for females.

For the International Youth Development Study (Hemphill et al., 2011), we have combined effect sizes on bullying (perpetration and victimization) versus theft for year 7 and year 10 students. Effect sizes for weapon carrying were also provided, but we felt that this fell more under violence and not offending behaviour. Follow-up periods were subsequently the same for bullying perpetration and victimization. The authors also controlled for the same number of covariates.

For the Japanese Longitudinal Study (Nishino 2010), results (for both bullying and victimization) are based on a combined effect size for shoplifting and vehicle theft. The same follow-up period and number of covariates controlled for are used for both predictors. The authors have provided via email communication zero-order correlation coefficients (unadjusted effect sizes) and standardized regression coefficients (for the adjusted effect sizes; they are shown in table 9) and sample size. These results are not presented in the subsequent special issue publication (Nishino et al., 2011). They have also provided, however, the additionally explained outcome variance after controlling for risk factors (i.e. the ΔR Square) from which we were able to obtain the partial correlation coefficient (which is what we have used in the meta-analyses). Preference was given to the most recent report over older ones (i.e. Nishino et al., 2009).

For the Metropolitan Area Child Studies (2 separate cohorts; see Henry et al., 2010) the authors provided unadjusted and adjusted effect sizes for bullying and victimization versus a total delinquency score based on correlation coefficients. The same covariates and follow-up periods are presented.

Some studies provided an effect size for bullying victimization versus offending only. This was the case with the Mater-University

of Queensland Study of Pregnancy and Its Outcomes, the National Longitudinal Survey of Youth 1997 and the Swiss Federal Survey of Army Recruits of 1997. For the Mater-University of Queensland Study of Pregnancy and Its Outcomes, McGee et al. (2011) show unadjusted and adjusted effect sizes for delinquency as an outcome measure. The authors show two adjusted effect sizes. We have chosen an effect size for victimization versus delinquency after controlling for family poverty and physical punishment and not the adjusted effect size after controlling also for age 5 aggression because we are interested in level analyses and not change analyses (and aggression is a construct very close to bullying). However, we should indicate that the two types of effect sizes were very similar in magnitude.

With regard to the National Longitudinal Survey of Youth 1997 (Wong, 2009), the author shows unadjusted and adjusted effect sizes for three different offences based on self-reports and we have combined the relevant effect sizes. With reference to the Swiss Federal Survey of Army Recruits of 1997 (Azzuzzi & Killias, 2010), the authors provided unadjusted effect sizes for victimization versus various outcomes. We have combined the effect sizes for knifed, strangled, shot with gun/stones and shot with firearm (4 separate measures)⁶⁹. This is the only retrospective longitudinal study on offending included in our meta-analyses. We have given preference to the Azzuzzi & Killias (2010) report over the older Haas (2001) report. We have not chosen to show results from the most recent report (i.e. Staubli & Killias, 2011) because the 2010 unpublished report actually presents more effect sizes.

Some studies provided an effect size for bullying perpetration versus offending only. This was the case with the Montreal Longitudinal Study, the Raising Healthy Children Project, the Seven Schools Longitudinal Study, the Pittsburgh Youth Study, the SNAP Under 12 Outreach Project and the Swedish Community Samples Study.

For the Montreal Longitudinal Study, an unadjusted OR for bullying perpetration versus delinquency could be computed from their table 6 (Tremblay & Haapasalo, 1998: 206). The paper also shows an adjusted effect size after controlling for family adversity (Tremblay & Haapasalo, 1998, table 5: 205) but the confidence interval is not provided. Eventually, we chose to report unadjusted and adjusted effect sizes from the Haapasalo et al. (2000) paper. We used the numbers in their table IV (Haapasalo et al., 2000: 155) to obtain an unadjusted OR. For adjusted effect sizes, we used the exponential of the regression coefficient from the logistic regression to obtain an OR (Haapasalo et al., 2000: 154) and we also used the value of the Wald statistic provided in order to compute

⁶⁹ The authors have also provided via email communication effect sizes for 'beaten/kicked', 'pulled a gun', 'beaten with heavy object' and 'chained'. We felt that these are more appropriate measures for violence and not offending.

the standard error and the relevant confidence intervals.

For the Raising Healthy Children Study (Kim et al., 2011), which involved a longitudinal preventive intervention to reduce problem behaviour, the authors have provided measures of association for bullying perpetration versus violence. The authors provide data indicating that the intervention did not confound the associations between childhood predictors and young adult problem behaviours. For calculating the unadjusted effect size, we used the zero-order correlations. These were estimated based on a covariance model in which maximum likelihood estimation was used to handle missing data and, therefore, all correlations are based on the same sample of 957 individuals (email correspondence with Dr Kim, October 30, 2010). For calculating the adjusted effect size, we used the standardized regression coefficients and treated them as partial correlation coefficients. For the Pittsburgh Youth Study (Farrington et al., 2011), the authors show unadjusted and adjusted effect sizes based on self-reports and mother-reports and for different age ranges. Their table 9 shows the combined (unadjusted and adjusted) effect sizes across the various age ranges (but separately for the two informants). The actual effect sizes for each age range are shown in the paper (Farrington et al., 2011, table 3: 78). In the final meta-analyses, we used a combined mother-rated and child-rated measure for both the unadjusted and adjusted effect sizes.

For the SNAP Under 12 Outreach Project unadjusted and adjusted odds ratios are provided for bullying perpetration versus offending based on official records (Jiang et al., 2011). Official record data in the form of odds ratios were also used for the Swedish Community Samples study (Olweus, 2011). Only unadjusted effect sizes were provided for this study. Finally, for the Seven Schools Study (Kendrick & Stattin, 2010), zero-order correlation coefficients were provided by Kendrick via email communication for the association of bullying (perpetration and victimization) versus property crimes based on self-reports. Kendrick has also provided adjusted effect sizes, but those were not used because, in the multivariate analyses, the author also controlled for property crimes at the baseline period; as already mentioned, for the aims of this review, we are not interested in analyses of change but only in analyses comparing levels of predictors with levels of outcomes.

Appendix 2: Combining Data on Depression Outcomes

For the Adolescent Mental Health Cohort Study, Kaltiala-Heino et al. (2010) show unadjusted and adjusted effect sizes separately for males and females. Same follow-up periods and covariates/confounders were used for bullying perpetration and victimization for both genders. For the Australian Temperament Project (Renda et al., 2011), results on the total sample were provided via email communication with Jenny Renda (see table 8). The same follow-up periods and covariates/confounders were used for bullying perpetration and victimization. The results were provided for the total sample.

For the Christchurch Health and Development Study, Gibb et al. (2011) provided effect sizes for bullying victimization versus depression based on parent reports in adolescence. Effect sizes for bullying perpetration in middle childhood versus depression were based in a combined parent-teacher report (the authors provided effect sizes separately for each informant as well). Effect sizes for bullying perpetration in early adolescence versus depression were based on parental reports only. For these reasons (which were also applicable for the two meta-analyses on offending outcomes) follow-up periods are different for perpetration and victimization. Different number on confounders have also been used for each predictor (see table 8).

For the Danish Longitudinal Retrospective Study (Lund et al., 2008) unadjusted and adjusted effect sizes were available for bullying victimization versus depression only. For the Dunedin Longitudinal Study (Moffitt et al., 2010), unadjusted and adjusted effect sizes were provided for bullying perpetration versus depression. Two outcome measures were provided: ‘ever diagnosed at age 32’ and ‘ever diagnosed at age 18 – 32’). In the relevant meta-analysis, we chose to include effect size data for ‘ever diagnosed at age 32’ because this provided the longest follow-up period. The results on bullying victimization were not provided.

Unadjusted and adjusted effect sizes for both predictors versus depression were provided via email communication with Susan McVie (2010) for the Edinburgh Study of Youth Transitions and Crime. Effect sizes are presented for the total sample. The same follow-up periods and number of confounders were used for the two predictors.

For the Erlangen-Nuremberg Development and Prevention Study (Lösel & Bender, 2011), a self-report measure of depression was used, with the same covariates and follow-up periods included in the analyses for both predictors. Results were shown separately for boys, girls and the total sample. We chose to report effect sizes based on the total sample. The same covariates and mean age at baseline and follow-up period was used for the association of bul-

lying (perpetration and victimization) versus depression for the Erlangen-Nuremberg Longitudinal Study of Bullying (Bender & Lösel, 2011; Lösel & Bliesener, 2003; Lösel et al., 2008), in which two measures of depression were available. We combined the relevant effect sizes. In the Lösel & Bliesener (2003) book, written in German, the results reported in our review are also presented along with other findings of the study. The Lösel et al (2008) paper at the European Society of Criminology conference presented the most up-to-date results, which are also shown in the Bender & Lösel (2011) manuscript. Across the three reports, there were no major differences in the effect sizes of interest. In both the above studies, the authors show a zero-order correlation coefficient as their unadjusted effect size. For ease of comparison with the correlation coefficients, the authors used the square root of the increase in R^2 in the second step of regression models (after controlling for covariates in the first step) as the adjusted measure of effect size.

The 'From a Boy to a Man' Finnish Longitudinal Study (Haavisto et al., 2004; Klomek et al., 2008; Sourander et al., 2007b) is part of the Nationwide 1981 Finnish Cohort Longitudinal Study (Sourander et al., 2009), but results are presented specifically for males. The Sourander et al. (2009) report presents results on the association of bullying (perpetration and victimization) with depression (more specifically, use of anti-depressants) on both males and females. We were unable to include effect sizes from this manuscript in our report because the authors used hazard ratios as their measure of effect size. The authors show numbers in the dichotomies for the unadjusted effect sizes (from which we would have been able to obtain an unadjusted OR), but they do not show numbers in the dichotomies for the adjusted effect sizes. Subsequently, we were not able to include this report in our meta-analysis on bullying perpetration/victimization versus depression. Instead, we were restricted in reporting effect sizes for the male population only. We have excluded the reports where only unadjusted effect sizes are presented (i.e. the Haavisto et al., 2004 and the Sourander et al., 2007b reports). Table 10 shows unadjusted and adjusted effect sizes for the Finnish longitudinal study based on the Klomek et al. (2008) paper, where ORs are shown for bullies, victims and bully-victims. We have combined bullies and bully-victims to obtain an overall effect size for bullying perpetration versus depression (unadjusted and adjusted). Similarly, we have combined victims and bully-victims to obtain an overall effect size for bullying victimization versus depression (unadjusted and adjusted). The only slight limitation of this paper is that adjusted effect sizes were shown after controlling for depression at a younger age ('change analyses'), but this was the best report available given the aims of our review.

For the Gatehouse Project, Bond et al. (2001) present unadjusted and adjusted effect sizes (in the form of ORs) for bullying victimiza-

tion versus depression based on the total sample. Results on bullying perpetration were not available. Pirkola and colleagues (2005) show unadjusted and adjusted effect sizes for bullying victimization at school (retrospective measure) versus adult health problems within the Health 2000 project. Unadjusted effect sizes are presented separately for males and females. The authors also show adjusted effect sizes separately for males and females after controlling for basic socio-demographic variables (their table 3: 733). However, they also present adjusted effect sizes (combined for gender) after controlling for other major risk factors such as maternal alcohol problems and paternal mental health problems (their table 4: 774). We have combined the separate effect sizes for males and females to get an overall unadjusted effect size for bullying victimization versus depression. Of the two types of adjusted effect sizes provided, we have chosen to report the second (combined for gender/ based on the total sample) adjusted effect size, since in these analyses important confounds (and not just socio-demographic variables) were controlled for.

For the International Youth Development Project two manuscripts were available (Hemphill et al., 2011; Patton et al., 2008). For the meta-analyses, we did not take into account the report by Patton et al., (2008) because: (a) only adjusted effect sizes for bullying victimization are given and, primarily, because: (b) a more recent study is available (Hemphill et al., 2011) in which both unadjusted and adjusted effect sizes for both predictors are presented. For this study, effect sizes were given separately for Year 7 and Year 10 students. We have combined the relevant effect sizes and reported (as with all studies where effect sizes from different groups were combined) an average age at Time 1, Time 2 and the follow-up period.

For the Japanese Longitudinal Study (Nishino et al., 2009; Nishino et al., 2011), we were initially able to locate the 2009 paper, which is written in the Japanese language (but with an English abstract). We got in touch with the first author and eventually obtained data relevant to the aims of our review. Through our correspondence, the Japanese team eventually prepared a paper for the special issue of *JACPR*. The two Nishino et al. (2009, 2011) papers are based on the same longitudinal study, with the most recent paper written specifically for the aims of our review and with the longest follow-up period⁷⁰. Unadjusted (zero-order correlation coefficients) and adjusted (standardized regression coefficients) effect

⁷⁰ The table with adjusted effect sizes in the Nishino et al. (2011) paper is wrong. The first author, who was also the corresponding author, was moving to a different university when proofs were due and mistakes in the proofreading of the paper subsequently occurred. The correct values are shown in this report. They were sent to us via email communication with Dr Nishino on the 26th of October 2010.

sizes for bullying victimization versus depression at three short-term follow-ups (covering a total span of about two years) were provided in the 2011 paper. We have combined the relevant values and reported an average age at follow-up for this study. The results were presented separately for males and females and we have combined them. For bullying perpetration versus depression, the results are based on one short-term follow up and are not gender-specific. These results were provided via email communication (see notes in tables 6, 7 and 8) with Dr Nishino and are not presented in the 2011 published paper. For the adjusted effect sizes on bullying perpetration versus depression, we have square-rooted the increase in R^2 when the predictor was entered in the second step (once covariates were controlled for/entered in the first step) in order to obtain the partial correlation coefficient.

Unadjusted (zero-order correlation coefficient; their table 2, on page 159) and adjusted (partial correlation coefficients; see page 158) effect sizes for bullying victimization versus depression are provided for the Longitudinal Retrospective Study of American University Students (Roth et al., 2002). Results were based on a 'teasing questionnaire', but no results were provided for bullying perpetration.

For the Mater-University of Queensland Study of Pregnancy and its Outcomes (McGee et al., 2011), unadjusted and adjusted effect sizes were provided in the form of ORs, for the total sample as well as for each gender separately. We have chosen to report the effect size for the total sample (their table 1: 112). Data were provided for bullying victimization only. Similarly, results on the Pittsburgh Youth Study were available for bullying victimization only. The authors (Farrington et al., 2011a) show unadjusted and adjusted effect sizes in the form of ORs based on self-reports and mother-reports and for different age ranges. Table 10 shows the combined (unadjusted and adjusted) effect sizes across the various age ranges (but separately for the two informants). The actual effect sizes for each age range are shown in the paper (Farrington et al., 2011: 79, table 4). In the final meta-analyses, we used a combined mother-rated and child-rated measure for both the unadjusted and adjusted effect sizes.

Results from two independent cohorts relating to the Metropolitan Area Child Study (Henry et al., 2010) were provided via email communication with Dr Henry (see notes in table 7). Unadjusted (zero-order correlation coefficients) and adjusted (partial correlation coefficients) effect sizes were provided on the association of bullying (perpetration and victimization) with depression for the total sample.

For the Seven Schools Longitudinal Study (Ozdemir & Stattin, 2011), Ozdemir has provided via email communication unadjusted effect sizes (zero-order correlation coefficients) for bullying (perpe-

tration and victimization) versus depression based on continuous measures (for both predictors and the outcome). Unadjusted effect sizes were provided for two follow-up periods. We have combined Time 1 and Time 2 effect sizes for (baseline) bullying perpetration versus depression; and accordingly for victimization. Subsequently, we report only one unadjusted effect size for each predictor versus the outcome, and accordingly we report the mean length of the follow-up period by obtaining an average across the two follow-ups. With regard to the adjusted effect sizes, in the published paper (Ozdemir and Stattin, 2011, table 1: 100) the authors present standardized regression coefficients for three separate categories, namely: bullies, victims and bully-victims. Standardized regression coefficients are presented in the paper for two follow-up periods.

Ozdemir has provided us with the equivalent unstandardized regression coefficients and the standard deviation of the dependent variable (i.e. depression) at each follow-up period. We have used the above information (along with the number of individuals in each of the four categories at baseline: bullies, victims, bully-victims and neither) to obtain Cohen's *d* (and the equivalent SE) for each follow-up period. Table 10 provides the relevant information which is not presented in the published report. For example, for bullies at Time 1, $d = .0722$ with an equivalent SE of $.13439$. We have combined Time 1 and Time 2 effect sizes for victims and bully-victims (four values of Cohen's *d*) to obtain an average effect size for bullying victimization at the baseline versus depression. Similarly, we have combined Time 1 and Time 2 effect sizes for bullies and bully-victims to obtain an average effect size for bullying perpetration at baseline versus depression.

For the Zurich Project on the Social Development of Children and Youth (z-proso), unadjusted and adjusted effect sizes for bullying victimization versus depression were provided (Averdijk et al., 2011; table 2: 106). For our meta-analysis, we used bullying victimization based on the 'variety score' versus the combined parent/teacher/child reports on anxiety/depression (unadjusted and 'level'). We did this because child attrition between waves 2 (baseline) and 4 (follow-up) was significantly related to bullying intensity ($OR = 1.14, p < .05$), so it did not seem correct to use the 'combined intensity/variety bullying score'. This information is not shown in the final version of the published paper, but it was available in the initial version of the paper that was sent to us via email communication from Margit Averdijk (June 16, 2010). All predictors and outcome measures were based on continuous variables.

Unadjusted effect sizes only were provided for twelve studies, namely: the Confident Kids Programme (Berry & Hunt, 2009), the Danish Longitudinal Health Behaviour Study (Due et al., 2009), the Dutch Anti-Bullying Programme (Fekkes et al., 2006), the Follow-Up Study in Canada (Vaillancourt et al., 2011), the Kiva An-

ti-Bullying Programme (Salmivalli, 2010), the Longitudinal Retrospective Study at the Mood Disorders Unit Outpatient Clinic in Sydney (Gladstone et al., 2006), the Longitudinal Retrospective Study of Adult Twin Pairs (Gladstone & Parker, 2006), the Longitudinal Retrospective Study of English GBQ men (Rivers, 1999, 2001; Rivers & Cowie, 2006), the Longitudinal Retrospective Study of Japanese University Students (Matsui et al., 1996), the Multimedia Violence Prevention Study (Espelage et al., 2001), the Six-Month Follow-Up Study in Canada (Shelley, 2009; Shelley & Craig, 2010) and the Swedish Community Samples Study (Olweus, 1993c, 1994b).

For the Confident Kids Programme (Berry & Hunt, 2009), Caroline Hunt has provided via email an unadjusted effect size for bullying victimization versus depression for the control group only. Results on bullying perpetration were not available in their dataset. For the Danish Longitudinal Health Behaviour Study (Due et al., 2009), only an unadjusted effect size for bullying victimization versus depression was available. We got in touch with the authors (see notes in table 7), but we were unable to obtain an adjusted effect size for this study. Similarly, for the Dutch Anti-bullying Programme (Fekkes et al., 2006) unadjusted effect sizes for bullying victimization versus depression for the control group were presented. Data on bullying perpetration as well as data on adjusted effect sizes were not available.

For the Follow-Up Study in Canada (Vaillancourt et al., 2011), a zero-order correlation coefficient is provided for bullying victimization versus depression (table 1: 195). The authors show effect sizes for Time 1 bullying victimization versus depression at two follow-up periods; we have chosen the longest follow-up period (Time 3 depression). The authors also show adjusted effect sizes based on path models, but the standardized regression coefficient for Time 3 depression is not given in the relevant table (i.e. figure 2: 195). For the Kiva Anti-Bullying Programme, Salmivalli (2010) has provided via email communication unadjusted effect sizes for bullying perpetration/victimization versus depression based on data for the control schools, which did not participate in the intervention study. Different effect sizes were provided for self-rated and peer-rated bullying and victimization. We have combined the relevant effect sizes.

For the Longitudinal Retrospective Study at the Mood Disorders Unit Outpatient Depression Clinic in Sydney, Australia (Gladstone et al., 2006), unadjusted effect sizes for bullying victimization versus depression were provided for the total sample based on two different measures on depression (a self-reported and a clinician-reported measure). We have combined the relevant effect sizes since the two measures on depression were not mutually exclusive. For this study, adjusted effect sizes were available for anxiety but not

for depression. For the Longitudinal Retrospective Study of Adult Twin Pairs (Gladstone & Parker, 2006), an unadjusted effect size for bullying victimization versus depression (and anxiety) was presented (zero-order correlation coefficient; see their table 2: 90). The authors also show results of a path analysis (figure 2: 91), with anxiety and depression being included in the model along with bullying victimization and inhibition. Path analysis involves simultaneous multiple regressions with path coefficients being equivalent to regression coefficients adjusted for other variables. In this case, the path coefficient from bullying to depression is similar to a regression coefficient after adjusting for inhibition and anxiety. Similarly, the path from bullying to social anxiety is equivalent to a regression coefficient after adjusting for inhibition and depression. However, because anxiety and depression were strongly correlated, we chose not to use the relevant path coefficients.

An unadjusted effect size for bullying victimization versus depression is presented for the Longitudinal Retrospective Study of English GBQ men (Rivers, 1999, 2001; Rivers & Cowie, 2006). Exactly the same F value of 14.08⁷¹ is presented across the three reports for GBQ men who were or were not bullied at school because of their sexual orientation, so there was no issue of ‘choosing’ the most appropriate/ representative effect size (because, for example, of different follow-up periods).

For the Longitudinal Retrospective Study of Japanese University Students, Matsui and colleagues (1996) show unadjusted effect sizes for bullying victimization versus depression (zero-order correlation coefficient, see their table 2 on page 717). The authors also provide an ‘adjusted’ effect size based on a step-wise regression model in which a combined score for depression at Time 1 and victimization at Time 1 were entered in the first step, and with their interaction entered in the second step (their table 3: 718). For obvious reasons, we were not able to use the results from the regression analyses. In the *JACPR* meta-analysis, we mistakenly indicate that we have included only adjusted effect sizes for this study but actually we only included the unadjusted effect sizes.

Espelage et al. (2001) present data on the stability of bullying behaviour based on a four-month follow-up (January 1995 to May 1995) and focus their analyses on 6th graders for whom bullying behaviour had increased. The authors show results for depression, anger and misconduct (all at Time 1) versus bullying behaviour at Time 2 for the 6th graders as a way to explain the increase in bullying behaviour among these sub-group of students (and not for the total sample, since the 7th and 8th graders showed stability in bullying behaviour). In email correspondence, Dorothy Espelage has provided us with the zero-order correlation coefficient for bullying

⁷¹ To obtain d , we have square-rooted the: $[F * (n_1 + n_2) / (n_1 * n_2)]$.

at Time 1 versus depression at Time 2 for the total sample (6th, 7th and 8th graders) as well as for the sixth graders. The study was part of a bullying prevention programme and the correlation coefficient provided was based on students from either (experimental or control) group. Ideally, we would have been interested in the correlation coefficient for the control group only. However, the authors indicate that the programme was not effective in altering bullying behaviour (Espelage et al., 2001: 414 & 419), so we have eventually included the results (for the total sample) provided via email in our meta-analysis.

Two reports were found with relevant data from the Six-Month Follow-Up Study in Canada (Shelley, 2009; Shelley & Craig, 2010). The published paper is based on the unpublished PhD of the first author and the same results were available in each report. Unadjusted effect sizes (zero-order correlation coefficients) for each predictor and the outcome were available separately for each gender. We have combined the relevant effect sizes and we give an overall effect size for the total sample. Adjusted effect sizes were not available for reasons explained before (see relevant discussion in part 2.3 on inclusion/exclusion criteria).

In two book chapters, Olweus (1993c, 1994b) presents data on the association between bullying victimization (based on teacher and peer reports) at grade 9 (age 16) and later depression at the age of 23 based on a follow-up study of Swedish men. Only unadjusted effect sizes are provided. Results on bullying perpetration versus depression were not available.

Adjusted effect sizes only were provided for two studies, the European TMR Network Project (Singer, 2002) and the SET Project (Kimber et al., 2008a, 2008b). In her unpublished PhD thesis, Singer (2002) shows the association of bullying victimization with depression. The data are based on a longitudinal retrospective study, showing current health indicators and how they are related to a retrospective measure of being bullied at primary and secondary school. We were able to show only adjusted effect sizes for this study. Singer (2002, table 24: 173) shows unadjusted effect sizes only for significant values, failing to report the non-significant zero-order correlation coefficients. For the adjusted effect sizes, Singer (2002) shows regression coefficients, which are similar but not the same as correlation coefficients. Nevertheless, she also shows the additional variance explained in the regression models once bullying victimization was entered in the model (seven separate regression models with a different confound included in each model). We have square-rooted the increase in R^2 to obtain the partial correlation coefficient.

Following two publications by Kimber et al. (2008, a & b) on social and emotional training in Swedish schools for the promotion of mental health, we contacted Birgitta Kimber and Rolf San-

dell informing them about our meta-analysis and asking for their contribution. The two published reports present data on the effectiveness of the SET study, a school-based intervention that is implemented using an experimental-control group design with before and after outcome measures. Kimber and Sandell have provided us with partial correlation coefficients between bullying victimization at Time 1 (May 2001) and various outcomes (including depression) at Time 2 (May 2002; 1-year follow-up) after controlling for gender, children's SES and the intervention factor. Ideally, we would like to show unadjusted (e.g. zero-order Pearson's correlation) and adjusted (e.g. partial correlation coefficients controlling for gender and SES) effect sizes based on the control condition only. However, because of high attrition (only 30 children at the follow-up; email correspondence with Rolf Sandell, dated March 24, 2010), the authors included measures for students of both the experimental and control condition and controlled for the impact of the intervention. Therefore, for the SET follow-up study we present adjusted effect sizes only. For depression, the authors actually included a measure of 'non-depressiveness'. Results were reverse-coded before inclusion in the meta-analysis. We should note that the authors have provided correlation coefficients for Time 3 as well (May 2003; 2-year follow-up). However, as already mentioned, because of the high attrition rates (in other words, due to the very small sample size), these correlation coefficients were considered meaningless and we therefore show results for the impact of bullying victimization at Time 1 on the outcomes at the shorter-term follow-up (Time 2). As for the sample size on which the partial correlations were based, we back-calculated it as: $N = DF - 4$ (since the authors were controlling for 3 covariates).

Other Reports in this Series

Closed-Circuit Television Surveillance and Crime Prevention

Brandon C. Welsh, David P. Farrington

A systematic review of the effects of closed circuit television surveillance based on 41 evaluations from five countries.

ISBN: 978-91-85664-79-5

Improved Street Lighting and Crime Prevention

David P. Farrington, Brandon C. Welsh

A systematic review of the effects on crime of improved street lighting based on 13 evaluations from the United States and United Kingdom.

ISBN: 978-91-85664-78-8

Effectiveness of Neighbourhood Watch In Reducing Crime

Trevor H. Bennett, Katy R. Holloway, David P. Farrington

A systematic review of the effects of neighbourhood watch based on 36 evaluations.

ISBN: 978-91-85664-91-7

The Influence of Mentoring on Reoffending

Darrick Jolliffe, David P. Farrington

A rapid evidence assessment of the effects of mentoring based on 18 evaluations.

ISBN: 978-91-85664-90-0

Effects of Early Family/Parent Training Programs on Antisocial Behavior and Delinquency

Alex R. Piquero, David P. Farrington, Brandon C. Welsh, Richard E. Tremblay, Wesley G. Jennings

A systematic review of effects of early family/parent training based on 55 evaluations from different parts of the world.

ISBN: 978-91-85664-95-5

Effectiveness of Programmes to Reduce School Bullying

Maria M. Ttofi, David P. Farrington, Anna C. Baldry

A systematic review of 59 scientific evaluations of anti-bullying programmes from different parts of the world.

ISBN: 978-91-86027-11-7

Effectiveness of Treatment in Reducing Drug-Related Crime

Katy R. Holloway, Trevor H. Bennett, David P. Farrington

A systematic review of the effects of drug treatment programmes on crime based on 75 evaluations.

ISBN: 978-91-86027-10-0

Effectiveness of Programs Designed to Improve Self-Control

Alex R. Piquero, Wesley G. Jennings, David P. Farrington

A systematic review of the effects of programs designed to improve self-control.

ISBN: 978-91-86027-38-4

Effectiveness of Interventions with Adult Male Violent Offenders

Darrick Jolliffe, David P. Farrington

A systematic review, including statistical meta-analysis, of the effects of programmes for preventing future offending of adult male violent offenders.

ISBN: 978-91-86027-39-1

Effectiveness of Public Area Surveillance for Crime Prevention: Security Guards, Place Managers, and Defensible Space

Brandon C. Welsh, David P. Farrington, Sean J. O'Dell

A systematic review of the effects on crime of three major forms of public area surveillance: Security guards, place managers and defensible space.

ISBN: 978-91-86027-49-0

Treatment Effectiveness in Secure Corrections of Serious (Violent or Chronic) Juvenile Offenders

Luz A. Morales, Vicente Garrido, Julio Sanchez-Meca

A systematic review and meta-analysis of 31 evaluations of the effectiveness of treatment programmes in secure corrections to prevent the recidivism of serious juvenile offenders.

ISBN: 978-91-86027-56-8

Preventing Repeat Victimization

Louise Grove, Graham Farrell, David P. Farrington and Shane D. Johnson

A systematic review of 31 evaluations of efforts to prevent repeat victimization, most focussing on burglary.

ISBN:

Bullying is a problem among children all over the world. In an earlier report in this series, two of the authors of this study have shown that systematic school programs have proven to be effective in preventing bullying. But what impact does bullying have on the bullies and those exposed to bullying later on in life? Does bullying have an impact on the risk for subsequent offending and mental health problems? What does the research tell us?

Systematic reviews are one means of helping people to pick their way through the jungle of research findings. Systematic reviews combine a number of studies that are considered to satisfy a list of empirical criteria for measuring effects as reliably as possible. The results of these studies are then used to calculate and produce an overall picture of the effects associated with a certain phenomenon. In this way systematic reviews systematically produce a more reliable overview based on the best well-founded knowledge available.

The Swedish National Council for Crime Prevention (Brå) has therefore initiated the publication of a series of systematic reviews, in the context of which distinguished researchers have been commissioned to perform the studies on our behalf. In this study, the authors have carried out a systematic review, including a meta-analysis, of 29 longitudinal studies of the impact of bullying on later offending, and of 49 longitudinal studies of its impact on depression.

David P. Farrington is Professor of Psychological Criminology at the Institute of Criminology, Cambridge University.

Professor *Friedrich Lösel* is Director of the Institute of Criminology, Cambridge University, and Professor of Psychology at the University of Erlangen-Nuremberg, Germany.

Dr. *Maria M. Ttofi* is Leverhulme and Newton trust Fellow at the Institute of Criminology, Cambridge University.

Nikos Theodorakis is a Ph.D. candidate at the Institute of Criminology, Cambridge University.



Brottsförebyggande rådet/National Council for Crime Prevention

BOX 1386/TEGNÉRGATAN 23, SE-111 93 STOCKHOLM, SWEDEN

TELEFON +46 (0)8 401 87 00 ▪ FAX +46 (0)8 411 90 75 ▪ E-POST INFO@BRA.SE ▪ www.bra.se

ISBN 978-91-86027-73-5 ▪ ISSN 1100-6676