Is Depression a Risk Factor for Adolescent Offending? A Meta-analytic Review

by

Dana Cochrane

B.A. (Hons.), Simon Fraser University, 2016

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Psychology
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Spring 2020

Approval

Name:	**Dana M. Cochrane**
Degree:	**Master of Arts (Psychology)**
Title:	**Is Depression a Risk Factor for Adolescent Offending? A Meta-analytic Review**
Examining Committee:	**Chair:** Matthew Sigal Lecturer
	Jodi Viljoen Senior Supervisor Professor
	Robert McMahon Supervisor Professor
	Jennifer Wong External Examiner Associate Professor School of Criminology
Date Defended/Approved:	April 17, 2020

Abstract

The literature is mixed about whether depression is a risk or protective factor for violence and general offending in adolescents. A meta-analytic review was conducted on 29 studies reporting on 27 unique prospective samples, with a total of 97,316 participants. The majority of samples were community (non-offender) or population samples (77.8%, k = 21), with a smaller proportion being justice-involved (e.g., incarcerated, probation, or history of arrest) samples (22.2%, k = 6). Overall, depression was associated with an increased risk for general offending (OR = 1.58, p < .001), and violent offending (OR = 1.45, p < .001). For community adolescents, depression was a significant risk factor for general offending; however, in justice-involved youth depression was not a significant risk factor. Gender, study quality, publication year, and country of publication did not moderate any of the results.

Keywords: depression; offending; violence; adolescence; delinquency; meta-analysis

Dedication

I would like to dedicate this to Mom and Dad. The biggest life lesson I've learned throughout the duration of this project is that 99% of the time your parent's advice is always right. Mom - you gave me my passion for the things that I love, and Dad – you gave me my insatiable curiosity and love of learning.

Acknowledgements

First, to my senior supervisor, Dr. Jodi Viljoen, thank you for your support, encouragement, and valuable mentorship throughout the years. I am extremely grateful for all the things you have taught me so far, and all the things I will learn in the future.

Thank you also to my committee members, Drs. Robert McMahon and Jennifer Wong for your insightful feedback that helped shape this project.

Thank you to the Social Sciences and Humanities Research Council of Canada for helping to fund this research.

I'd like to thank Dr. Nicole Muir for taking me under her wing when I was an undergrad student and encouraging me to pursue my goals, and Dr. Martin Davidson for being such an inspiring teacher that made me realize my passion for this field.

I'd like to thank Cindy for assisting me with this project, and Andrew for all his meta-analysis know-how. I'd like to thank my cohort and lab mates, who continue to make this journey so memorable.

To my parents, Keith and Connie, who encouraged my love for learning, and my sister, Nicole, I would not be who or where I am today without your love and support.

Last, but certainly not least, thank you to my partner who has been my unwavering rock during this journey. You support me in so many ways I cannot thank you enough.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

During the transition from childhood to adolescence, rates of mental illness, suicide, and emotional problems increase (Reiter, Suzuki, O'Doherty, Li, & Eppinger, 2019). Furthermore, recent studies have noted a rise in adolescent-onset mental illness; for instance, between 2000 and 2014, mood disorder diagnoses in Canadian adolescents increased significantly (Wiens et al., 2017). However, rising rates for adolescent mental illness is not only a concern for general samples of youth in the community, but also adolescents involved in the justice system. In particular, a recent national survey found that from 2004 to 2016 the past-year incidence rate of major depression has been increasing for both adolescent females and female juvenile offenders, with current rates jumping from 12% to 17% for non-offenders, and 24% to 33% for juvenile offenders (Holzer, Oh, Salas-Wright, Vaughn, & Landess, 2018). Despite the increased prevalence rates in offenders, young offenders in particular are less likely to have their mental health problems identified or treated than non-offending youth (Kenny, Nelson, & Lennings, 2007).

It has been well established that depressive symptoms and offending behaviour co-occur, particularly in adolescence, where onset of both problems typically emerge (Wiesner & Kim, 2006). Although depression may first appear at any age, typically the age of onset is around puberty, with the incidence rate peaking in the 20's (Kessler et al., 2003) and then leveling off in later adulthood. Similarly, one of the most well-accepted tenets in criminology is the age-crime curve (Fagan, & Western, 2005). Relatively few individuals display delinquent behaviours in childhood; however, involvement in delinquency rises throughout adolescence hitting a peak in late adolescence, and then declining in early adulthood (Fagan & Western, 2005). However, despite the fact that both problems tend to emerge at the same time, researchers have concluded that the degree of comorbidity between depression and offending exceeds what would be expected by chance (Loeber & Keenan, 1994).

The high prevalence of depression in juvenile offenders raises questions about the role that depression plays in offending. For instance, is this elevated rate of

depression in offenders spurious and simply due to shared risk factors, such as social disadvantage (Burt, 2009)? Or does depression serve as a risk factor for future offending and play a potentially causal role in offending? Thus far, theory and research are mixed. Some studies have found significant associations between depression and both general and violent offending (Boots, & Wareham, 2010; Blitstein, Murray, Lytle, Birnbaum, & Perry, 2005; Connell, & Dishion, 2006; Felson, Silver, & Remster 2012), whereas others have found nonsignificant results (Sigfusdottir, Farkas, & Silver, 2004), or even that depression might be protective (Farrington, Gallagher, Morley, & St. Ledger, 1988). These mixed results could mean the impact of depression might vary by gender or by other characteristics such as offence type. However, at this point these potential moderators have not yet been adequately explored.

As such, the goal of the present study is to empirically synthesize longitudinal research on whether depression predicts future offending in adolescence. Rather than simply looking at main effects, the current study examines potential moderating variables, particularly gender, but also factors such as population (e.g., school students vs. juvenile offenders on probation). Prior to describing this study, I review theory on the relationship between depression and offending, as well as prior research. Then, I discuss the importance of studying depression and offending in adolescence from a developmental perspective. Lastly, I review factors that might moderate the association between depression and offending, such as gender and the type of offending.

Theoretical Models of Depression and Offending

Two main programs of research have examined the relationship between depression and offending: mental health research and criminological research. The theories posed by mental health research comes from a sense of urgency regarding the high prevalence rates of depression and other disorders in offender populations and wanting to find a solution to help mentally ill offenders (McCormick, Peterson-Badali, & Skilling, 2015). Mental health research presents four main frameworks to conceptualize the relationship between depression and offending: shared risk factors, the acting-out model, the behavioural inhibition model, and the failure model.

Shared Risk Factors Model. According to the shared risk factors model, depression and offending co-occur frequently (Wolff & Ollendick, 2006). However, their

co-occurrence is caused by nonspecific risk factors that collectively lead to separate but associated problems (Wolff & Ollendick, 2006). Some examples of potential shared risk factors are poverty, childhood maltreatment, and substance use (Wolff & Ollendick, 2006).

Acting Out Model. In contrast to the shared risk factor model which simply posits that depression and offending are simply spuriously correlated with each other, the acting-out model argues that offending behaviour is caused by depression. From this perspective, depressive symptoms, particularly irritability, may be expressed behaviourally through heightened rule breaking and aggression (Aske, Hale, Engels, Raaijmakers, & Meeus, 2007; Wolff & Ollendick, 2006). Over time, the repeated expression of this emotional distress via acting out will lead to serious offending behaviour.

Behavioural Inhibition Model. Whereas the acting-out model hypothesizes that depression increases offending, the behavioural inhibition theory hypothesizes that depression decreases offending (Kerr et al., 1997). The basic premise of this model is that certain symptoms of depression, such as apathy and reduced energy, prevent behaviour such as spending time with antisocial peers that would make them susceptible to getting involved in offending behaviour (Hein et al., 2017).

Failure Model. Contrasting with the acting out model and behavioural inhibition model that conceptualize depression as either increasing or decreasing the risk of delinquency, the failure model assumes that delinquency causes subsequent depression (Capaldi, 1992). Engaging in delinquent behaviours may result in rejection and failure in social relationships as well as academic failure. This lack of normative success place children and adolescents at an increased risk for depression (Capaldi, 1992).

In addition to these mental health theories that focus on community adolescents, several criminological theories (i.e., the Risk-Need-Responsivity model) have attempted to understand the relationship between depression and offending in people who are already part of the criminal justice system.

RNR Model. According to the Risk-Need-Responsivity (RNR) model, which is the dominant model of offender rehabilitation (Bonta & Andrews, 2016), depression has a “very minor or no causal relationship to criminal behaviour” among individuals in the

justice system (Andrews & Bonta, 2010, p. 45). The need principle of the RNR model dictates that offenders should receive treatment that targets their specific criminogenic needs. Criminogenic needs are dynamic risk factors, such as substance use or antisocial attitudes, that when reduced will decrease risk for recidivism (Andrews & Bonta, 2010). The RNR model views depression to be a non-criminogenic need, meaning that it is viewed as generally not relevant to re-offense risk.

Instead, within the RNR model, mental disorders such depression are considered a responsivity factor. Andrews and Bonta (2010) describe responsivity factors as "personal characteristics that regulate an individual's ability and motivation to learn" (p.46). These individual factors would be considered anything that needs to be tailored to ensure full participation in treatment efforts. If depression acts as a barrier to engaging in treatment, mental health professionals should tailor interventions so that an offender's depression does not interfere with treatment. However, this perspective leaves depression as a side consideration rather than a focal point. Some authors argue that this is insufficient. For instance, according to McCormick, Peterson-Badali, and Skilling, (2017), the justice system should not only aim to reduce recidivism, but also to reduce suffering.

Strain Theory. Whereas in the RNR model depression is not a central focus, General Strain Theory (GST) places a larger emphasis on the role that negative emotions play in offending. Although not developed specifically to capture depression per se, GST is based on the premise that strains and stress trigger negative emotions like anger and frustration if someone has a lack of coping skills (Agnew, 1992). In this perspective, Agnew posited that a depressed person sees committing crime as a potential solution to reduce emotional pressure, just as someone with depression may also use substances to cope with negative emotions (Agnew, 2006). However, GST does not clarify whether if depression was found to be positively correlated with offending, would this indicate depression as a cause of offending, or perhaps that a third variable such as anomie accounts for both depression and offending? That is, depression may be a risk marker for offending, but not a causal risk factor (Agnew, 1992).

In sum, it is clear that different theories conceptualize the relationship between depression and offending in different ways. Below, I review research on the extent to which these differing hypotheses have empirical support.

Empirical Findings

In general, researchers have reported mixed findings on the association between depression and offending in adolescents. Some studies have found depression to be a clear risk factor for general offending and violence (Blitstein, Murray, Lytle, Birnbaum, & Perry, 2005; Boots, & Wareham, 2010; Connell & Dishion, 2006; Felson, Silver, & Remster 2012; Ritakallio, Kaltiala-Heino, Kivivuori, & Rimpelä, 2005), and some studies have found no relationship (Sigfusdottir, Farkas, & Silver, 2004). In other cases depression has been found to be a protective factor against offending (Farrington, Gallagher, Morley, & St. Ledger, 1988; Kerr, Tremblay, Pagani, & Vitaro, 1997; Pfeffer, Plutchik, & Mizruchi, 1983).

Overall, the link between depression and violent offending in adolescence has a stronger empirical foundation than the link to general offending. For example, Yu et al. (2017) used three separate longitudinal datasets to test to see whether the depression-violence link held for adolescents. Depression predicted a 1.7 to 2.8 times increased risk for subsequent violence. However, in each of the three datasets, general offending was not analyzed as an outcome, only violent offending. So, what might appear in the literature as a more solid foundation for the link between depression and violence, may just be an artifact of the dearth of high-quality longitudinal research examining general offending as an outcome. This may be a consequence of researchers' viewing the link between depression and violence as more plausible than depression and general offending. General offending is extremely broad and by definition encapsulates all types of offending ranging from probation violations, property crime, drug offences, and violent offences. This makes it more difficult to conceptualize how depression may impact all these types of offending, when it very may well be that depression differently impacts the different types of offending

Although meta-analysis is a useful way to synthesize and make sense of differing findings, to date, only two meta-analyses have examined the relationship between depression and offending, and these studies focused primarily on adult offenders. As

part of a larger meta-analysis on predictors of recidivism, Bonta et al. (1998) synthesized nine studies on adult offenders, and failed to find a significant association between mood disorders and violent or general recidivism. As a follow up Bonta, et al. (2014) examined 13 studies on the relationship between mood disorders and offending. Again, this meta-analysis found no relationship between mood disorders and general and violent recidivism. However, several important gaps in knowledge remain.

First, although those meta-analyses focused on adults, depression may manifest differently for adolescents and adults and thus the effect it plays on offending may differ between these developmental periods. According to the concept of age relativity, age determines what kinds of behaviours and emotions we can identify as symptoms of depression (Mash & Dozois, 2003). For instance, difficulties going to bed and falling asleep may not be a sign of depression in children and adolescents given normative issues with sleep hygiene in younger individuals; however, this may be a symptom of depression in adults. Similarly, whereas depressed mood in adults is a hallmark sign of depression, the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5; American Psychiatric Association, 2013) recognizes that this typically manifests as irritable mood in children and adolescents. As a consequence of the age relativity of disorders, researchers argue that characteristics can only be considered to be a symptom of a disorder if they diverge from the "average" behaviour of individuals of the same developmental period and result in impairment (Cicchetti & Rogosch, 2002). Thus, depression in adults and youth needs to be understood as related, but separate entities with differential manifestations.

Second, the meta-analyses only included studies that sampled offenders receiving mental health treatment, rather people with no previous history of justice involvement. As acknowledged in developmental psychopathology and developmental criminology research, the factors that predict first incidence of a negative outcome may differ from those factors that precipitate subsequent occurrences; thus, both initial onset and maintenance of offending should be examined (Fagan & Western, 2005; Rutter & Sroufe, 2000). For instance, although depression may not be a risk factor for recidivism in people who are currently involved in the justice system, it may predict onset of offending for normative youth.

Third, these meta-analyses did not examine depression as a stand-alone variable and rather combined depression into a general mood disorders category along with bipolar disorder; as such, it remains unclear how depression is connected to offending. Lastly, moderators of the relationship between mood disorders and recidivism were not examined in either meta-analyses. Due to the many ways in which depression can present, and the numerous individual variables that may interact with depression, moderating variables between depression and offending should be evaluated to further elucidate why the findings are mixed. For instance, the relationship between depression and offending might differ for male and female youth.

Compared to males, females engage in lower levels of offending behaviour during adolescence and have higher rates of depression (Holzer et al., 2018). However, despite depression being more common in female adolescents, some researchers have observed that that the strength of association between depression and offending is stronger in males than females (Kim & Kim, 2005; Obeidallah, Brennan, Brooks-Gunn, & Earls, 2004; Ritakallio et al., 2008). One explanation may be the gender-specific risk enhancement theory (Taylor & Ounsted, 1972). This theory argues that, as boys typically show lower prevalence rates of depression than girls, the boys who are depressed will experience depression more severely (Loeber & Keenan, 1994). In other words, because depression in males is rarer, when a male is depressed, it has more negative consequences than it would for females. Another explanation for why depression may be a stronger predictor of offending in males than females is that depression in males may be more stigmatized (Shaffer, 1998). If depressed males express their emotional distress in a gender-atypical way, this could lead to more peer rejection than depressed females receive which could, in turn, place them at higher risk to offend (Shaffer, 1998).

The Current Study

In sum, given the complexity and challenges posed by the very nature of the co-occurrence of depression and offending and similar developmental trajectories, the area is left with disparate theories and empirical findings. Some researchers hypothesize that depression is positively associated with offending specifically by symptoms such as irritability being ‘acted out’ in offending behaviour. Other scholars believe that depression is negatively associated with offending because symptoms like apathy and fatigue inhibit offending behaviour. Still others view that delinquent behaviour leads to

multiple failures which causes subsequent depression. Lastly, other researchers view depression as unrelated to recidivism entirely. As such, it is currently unclear which conclusions are justified. This results in a lack of empirical evidence to inform policy and practice on whether depression should be prioritized as a focal point of rehabilitative efforts or whether money and efforts can best be used more efficiently elsewhere. Thus, I aimed to synthesize research to help disentangle the relationship between depression and offending in adolescence.

The goal of the present meta-analysis was to expand on the previous meta-analyses by Bonta and colleagues (1998, 2014). First, rather than collapsing bipolar and depression into one variable, I included variables that assess depression alone so as to not conflate the findings with other disorders. Second, instead of focusing exclusively on offender populations, both general community samples and samples of justice-involved adolescents were included to determine if there are population differences. Third, the type of offending as an outcome was not restricted; violence, general offending, and intimate partner violence were assessed as outcomes. Fourth, I examined potential moderating variables such as gender, type of offending, population type (community sample vs. justice-involved youth) and follow-up length to help explain the heterogeneity in the findings. The current study aimed to clarify for whom depression is a relevant risk factor, for which types of offences, and which factors may moderate the relationship.

As such the research questions were as follows:

1. What is the overall relationship between depression and offending?
2. Does the relationship between depression and offending vary between community and justice-involved adolescents?
3. Does the relationship between depression and offending vary between males and females?
4. Which other factors moderate the effect of depression on offending?

Chapter 2.

Method

To address the research questions, a meta-analytic review was conducted. The guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) were followed to ensure that the review was rigorous, transparent, and as thorough as possible. For example, specific search terms, databases, and dates of searches were reported to allow the search to be replicated. An initial literature review was conducted before data collection commenced to ensure enough studies existed on the topic for a meta-analysis to be appropriate, and to help inform the development of research questions and the data extraction form.

Step 1: Search Methods

The search procedure is outlined in Figure 1. To select the search terms, seven combinations of terms were pilot tested by recording how many studies were identified by each combination, and whether each combination was able to capture studies already known to meet inclusion criteria. The final set of search terms was chosen as they were specific enough to identify studies that met inclusion criteria, but at the same time, was not too broad to identify an unwieldy number of studies (e.g., 20,000). The final set of search terms were as follows: depress* AND (criminal* OR devian* OR violen* OR delinquen* OR offend* OR offense* OR offence* OR reoffend* OR recidiv*) AND (longitudinal OR follow up OR time points OR waves OR prospective) AND (adolescen* OR youth OR juvenile OR teenager). Using these terms, 13 databases were searched, including the following: PsycINFO, PsycARTICLES, PsycBOOKS, MEDLINE, National Criminal Justice Reference Service, Criminal Justice Abstracts, Sociological Abstracts, Social Sciences Abstracts, Social Sciences Full Text, Social Services Abstracts, and Web of Science.

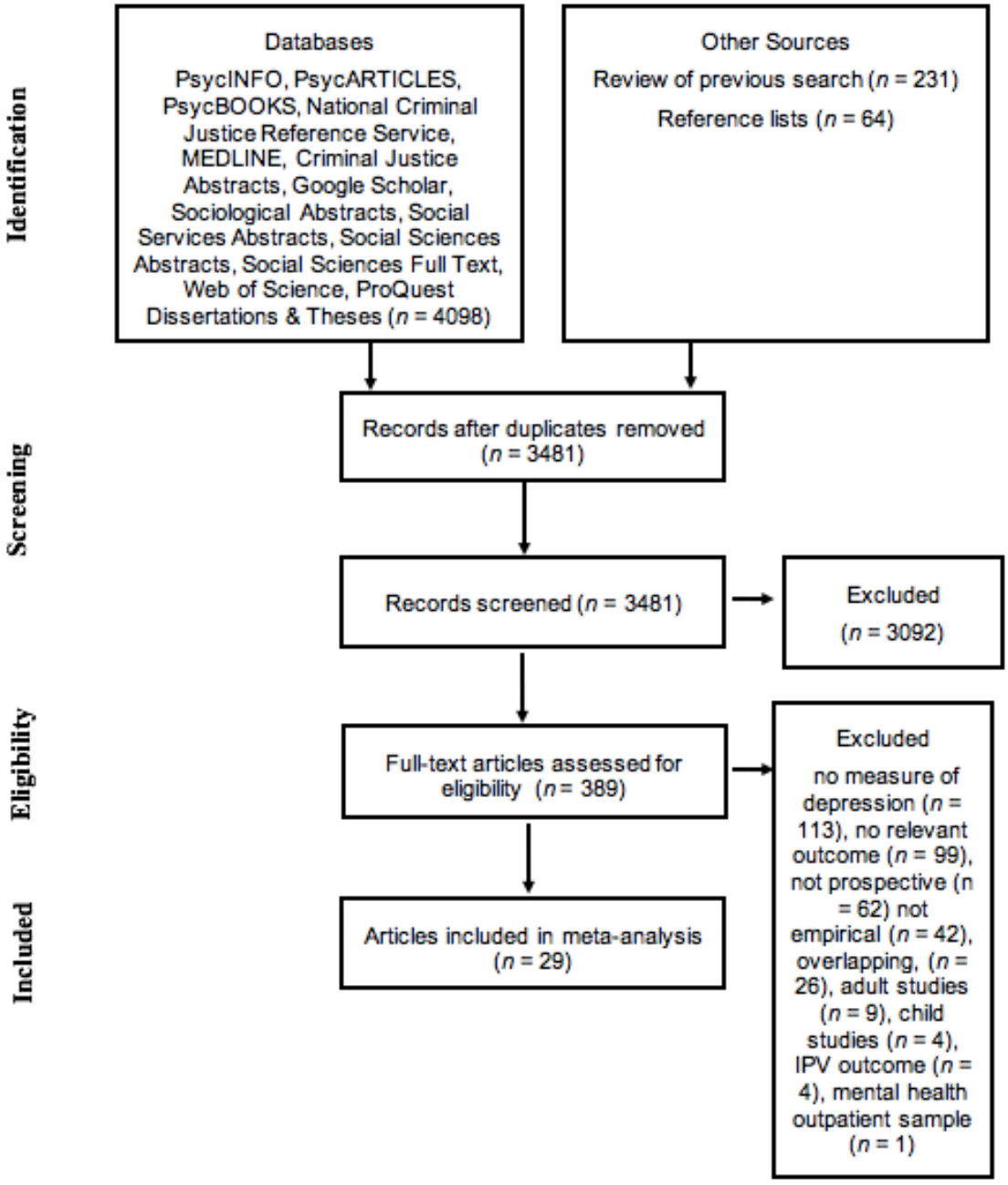

Figure 1 Search Strategy and Phases of Meta-analysis

To reduce the likelihood that publication bias affected the results of the review (Kicinski, Springate, & Kontopantelis, 2015; Loannidis, 2005;), other databases that capture more of the grey literature (ProQuest Dissertations and Theses Global Database, Google Scholar) were also searched. These searches encompass the year 1980 (which marks the publication of DSM-III [American Psychiatric Association, 1980],

which presented the first consensus-based description of depression) up to July 11, 2018, the day the searches were conducted. Often researchers restrict Google Scholar searches to 100 records (Haddaway, Collins, Coughlin, & Kirk, 2015); however, the first 300 Google Scholar search records were examined in the present review to increase the thoroughness of the literature search. In addition to searching databases, reference lists of included studies and the Bonta and colleagues' (1998, 2014) meta-analyses were also reviewed for relevant studies. In addition, the abstracts of a previous unpublished literature search on a related topic were reviewed (Viljoen et al., 2016).

Step 2: Abstract Screening and Inclusion Criteria

After removing duplicate articles, 3,481 studies were identified. The study author provided 2 hours of didactic training to the research assistant (RA), a fourth-year psychology undergraduate student, regarding the abstract screening procedure. Then, the RA and study author screened the first 100 abstracts independently and met to review decisions to ensure that all studies meeting inclusion criteria were identified to create a valid and reliable process. The rest of the abstract screening was divided equally between the RA and author. Studies were required to meet the following inclusion criteria: (1) empirical study with data presented in a manner to be able to calculate odds ratios, or to estimate from other statistics such as Cohen's *d*, or Pearson's *r*; (2) design was prospective; (3) included a measure of depression; (4) included an offending outcome relevant to a research question (general offending, violence) (5) included a sample of adolescents.[1] This study focused only on prospective designs due to the importance of establishing time-ordered relationships between depression and offending. If depression and offending are measured concurrently, then there is no way to establish if depression precedes the onset of offending. The current review only included studies in which participants' age at baseline was between 10 and 19 years (World Health Organization, 1986). Studies were excluded if: (1) the measure

[1] Although initial focus of the review was on both adult and adolescent samples, it was restricted to adolescent studies only because the adult samples were predominantly justice-involved (77.8%, *k* = 7), whereas the adolescent samples were predominantly community samples (77.8%. *k* = 21). Due to the differences in sample type, it made it difficult to meaningfully compare across groups. Therefore, it was decided that the present Masters thesis would only report and discuss results of the adolescent samples, and the adult samples would be reported in a separate study.

of depression was confounded with another disorder such as anxiety or bipolar; (2) the outcome would not be considered an offence (i.e., bullying/aggressive play).

Step 3: Full Text Review

The study author conducted a full text review of the 389 abstracts that were screened in. Of the articles, 29 studies met inclusion criteria (see Figure 1). The other articles did not meet the inclusion criteria for various reasons. In 113 articles there was no measure of depression, or the measurement confounded other disorders such as collapsing depression and anxiety together. In 99 articles there was no measure of offending. Sixty-two studies were not prospective and 42 studies were not empirical and therefore did not report any effect sizes that could be compiled. Twenty-six studies were excluded because they had samples that overlapped with other included studies. When disseminations were based on the same sample and measures, the study that was most comprehensive (e.g., largest *n*) was chosen. Regarding age, 9 studies were excluded because the age at baseline was greater than 19 years old and 4 studies were excluded because the age at baseline was younger than 10 years old. Finally, 4 studies were excluded that only reported outcomes of intimate partner violence and 1 study that sampled an outpatient mental health population was excluded due lack of congruence with the community samples and justice-involved samples.

Step 4: Data Extraction

Data from the studies that met inclusion criteria was extracted using a data extraction form (see Appendix A). This form was created by reviewing other data extraction forms (e.g., Morgan et al., 2016). The form was pilot tested with five articles and revised accordingly. A random sample of 10 studies was selected to be coded by a second coder for the purposes of calculating inter-rater reliability.

Sample Characteristics. Five variables were coded relating to the study sample. Sample setting was coded according to whether the population was justice-involved (e.g., probationers, incarcerated, arrestees), psychiatric patients, or whether it was a community or population sample (e.g., school students, national census). Gender was coded based on whether the sample comprised of entirely men, entirely women, or mixed gender. Country of data collection was coded, and publication status was coded

based on whether it was a peer reviewed published article, or part of grey literature including dissertations. The age of the sample was coded according to whether it was an adolescent, adult, or mixed sample. Cohen's kappa coefficients indicated perfect agreement (κ = 1.0) for sample setting, gender, country, and publication status and fell in the almost perfect range for age (κ = .81; Landis & Koch, 1977). Finally, sample size used in analyses was recorded, with an intraclass coefficient (two-way mixed, absolute agreement, single measure) of 1.00 indicating perfect agreement (Cicchetti, 1994).

Measurement of Depression. The method for assessing depression was coded as either a self-report measure that captures depression (e.g., Beck Depression Inventory II: BDI-II; Beck, Steer, & Brown, 1996), or official clinician diagnosis (e.g., Diagnostic and Statistic Manual of Mental Disorder, Fifth Edition: DSM-5; American Psychiatric Association, 2013). Agreement was perfect (κ = 1.0).

Measurement of Offending. Method of assessing offending was coded as: official records (such as official criminal justice records, hospital incident reports, or direct observation), self-report, and mixed/combined methods. Types of offending were categorized into general offending (which captures all types of offences), violent offending, and intimate partner violence. Agreement was perfect (κ = 1.0) for both variables.

Study Quality Measure. Study quality was rated using a modified version of the Newcastle – Ottawa Scale (NOS; Wells et al., N.D.) for cohort studies, which is a tool for assessing the quality of non-randomized studies in meta-analyses and systematic reviews (see Appendix B). This scale was chosen because it is a validated tool that uniquely assesses moderator variables, which are often not captured in other tools (Luchini, Stubbs, Solmi, & Veronese, 2017). Further the tool itself is intended to be adaptable to the research context in which it is being employed. Additionally, it is one of the most commonly used tools for evaluating study quality in observational studies (Luchini et al., 2017).

The NOS contains eight items that encompass three dimensions of study quality. The selection of participants domain captures the representativeness of the study participants and the quality of measurement of the predictor variable (i.e., depression in the current review). It contains four items: representativeness of exposed cohort, equal

derivation of exposed and nonexposed cohort, measurement of exposure, and demonstration that the outcome of interest was not present at the beginning of the study. The comparability dimension contains one item and assesses the extent to which the study controls for potential confounding variables in its analyses. The outcome dimension captures the quality of the measurement of outcome variable (i.e., offending in the current review), as well as the adequacy of follow-up length and the retention rate of follow up data. It contains three items (adequate assessment of outcome, adequate follow-up time, and adequacy of follow-up).

The tool was modified to make it more applicable to the studies in the current meta-analysis by removing an item that assesses whether the outcome of interest is not present at the beginning of the study. This item is relevant if assessing an outcome like a disease and what predicts the onset of the disease. For example, if assessing risk factors for cancer, it is important for cancer to be absent at the beginning of the study to ensure the temporal order of events is clear. Otherwise the risk factor being assessed (for example, poor sleep quality), may be a consequence of cancer rather than a risk factor that precedes the onset of cancer. However, the outcome of interest in the present study, offending, is expected to be present at the beginning of some studies (such as studies sampling incarcerated offenders), and thus questions pertaining to the absence of the outcome of interest are not relevant for the current study.

A study can be awarded a maximum of 1 point per item for all items with the exception that the comparability item can be awarded a maximum of 2 points. The maximum score on the revised Newcastle-Ottawa Scale is 8 (highest quality), with scores of 0-2, 3-5, and 6-8 for low, moderate, and high quality, respectively (Stang, 2010). The ICC (two-way mixed, absolute agreement, single measure) fell in the excellent range for the total score (.81; see Table 2 for individual study ratings; Cicchetti, 1994).

Data Analytic Strategy

Effect Sizes and Weighting Procedures. The first goal of the current study was to characterize the relationship between depression and different types of offending across all relevant studies. As such, weighted effect sizes were calculated in Comprehensive Meta-Analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2013)

software. Odds ratios were chosen as the effect size to report as the majority of studies included in the review (*k* = 19, 65.5%) used odds ratios to describe the relationship between depression and offending. Further, the use of odds ratios in meta analytic reviews have been recommended because they are directly comparable across studies with different designs (Borenstein et al., 2009). Additionally, odds ratios are fairly easy to interpret with the value indicating the increased relative risk of the outcome (offending) associated with the presence of the other variable (depression). Odds ratios of 1 indicate no relationship between the two variables, values greater than 1 indicate a positive association or increased risk, and values less than 1 indicate a negative relationship or decreased risk (Haddock, Rindskopf, & Shadish, 1998). In epidemiology, odds ratios of 2.0 – 2.5 and above represent meaningful associations, and odds ratios of 3.0 and above are considered to be large associations (Haddock et al., 1998).

Weighting in the current study was conducted according to the guidelines of Borenstein et al. (2009). Meta-analyses either use a fixed-effect or random-effects statistical model to combine and weight studies. As opposed to fixed-effects models, which assume that each study is estimating the same (or common) underlying effect, a random-effects model assumes that each study is estimating a study-specific effect (Borenstein et al., 2009). Under a fixed-effects model, heterogeneity across study effects is attributed to random sampling error, and therefore, hypothetically, if all studies had an infinite sample size, there would be no differences due to chance, and differences in study effects would disappear (Borenstein et al., 2009). In contrast, a random-effects model assumes that the effects vary across studies because of real differences in the effect of the predictor on the outcome as well as sampling variability (chance). Consequently, if each study hypothetically had an infinite sample size, differences would still exist across study effects due to the true differences in the effect of predictor (i.e., depression) and outcome (i.e., offending). These differences in effects can be caused by differences in study population, such as age, follow-up length, and other factors. Random effects models were employed as given the mixed findings in the literature regarding depression and offending, the true effect size will most likely differ between studies (Borenstein et al., 2009). For a random effects model, each study's effect size is weighted by the inverse of its variance plus the estimate of between-studies variance. To examine heterogeneity between study effect sizes, a within-group *Q* statistic tests the presence or absence of heterogeneity. Further, Higgins I^2 is an

indication of the proportion of variance due to heterogeneity. An I^2 < 25% is interpreted as low, 50% is medium, and 75% is high. (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006).

Moderation Analyses. An additional goal of the current study was to identify variables relating to study design, variable measurement, and sample characteristics that may moderate the relationship between depression and offending. There are two standard ways to analyze moderating variables in a meta-analysis: subgroup analysis and meta-regression (Borenstein et al., 2009). Subgroup analysis is the one-way analysis of variance analogue for meta-analyses, and meta-regression is a regression analogue designed to estimate regression coefficients for moderating variables (Borenstein et al., 2009). The advantage of meta-regression is that it can analyze both continuous and non-continuous types of variables. However, meta-regression requires a minimum of 10 studies per moderating variable, and if say 10 moderators were to be analyzed, this would require 100 studies for adequate power (Borenstein et al., 2009). Therefore, given that 29 studies were included in the current review, subgroup analysis was used for the majority of moderator analyses, and meta-regression was reserved for testing a few select moderators.

Chapter 3.

Results

Description of Included Studies

In total, this meta-analysis included data from 27 unique prospective samples, stemming from 29 manuscripts, with an aggregated sample size of 97,316 participants (see Table 1). Four separate samples were reported on twice in the present meta-analysis because the articles included different outcomes. For example, both Anderson (2015) and Siennick (2007) reported on the National Longitudinal Study of Adolescence to Adulthood Health (ADD Health) study. Anderson (2005) reported outcomes for violent offending, and Siennick (2007) reported outcomes for any offending. In addition, Yu et al. (2017) reported on three separate samples.

The majority of these manuscripts (86.2%, *k* = 25) were published in peer reviewed journals, while the remaining four (13.8%) were unpublished dissertations or theses. Most samples were collected in the United States (59.3%, *k* = 16), with the remaining samples from Europe (29.6%, *k* = 8), South Korea (7.4%, *k* = 2),and Canada (3.7%, *k* = 1). The majority of studies sampled community (non-offender) or population-based participants (77.8%, *k* = 21), with a smaller proportion of studies sampling justice-involved (e.g., incarcerated, probation, or history of arrest) adolescents 22.2%, *k* = 6).

Depression was measured in the 27 samples with self-report questionnaires (77.8%, *k* = 21), clinician-administered diagnostic interviews (*k* = 3, 11.1%), informant report (e.g., parental or teacher report; *k* = 2, 7.4%), and one study using official medical records (*k* = 1, 3.7%). Offending was assessed by self-report questionnaires in 70.4% of articles (*k* = 19), structured interviews (7.4% *k* = 2), official records (14.8%, *k* = 4), and combinations of self-report and official records (7.4%, *k* = 2).

Table 1 Studies Included in Meta-analysis

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 1	Anderson, 2015 (USA)	15,584 community adolescents (M/F)	ADD Health Waves 1 and 4 (approximately 13 years)	Center for Epidemiological Studies – Depression Scale (CES-D; Radloff, 1997; 18 out of 20 items, summed, and rescaled to match original 20 items); depression determined to be present by gender-specific cut off scores (in grades 7 through 12)	Self-report questions from ADD Health measure past year violent offending, converted to binary variable (aged between 25 and 32)	No	7
Article ID 2	Aske et al., 2007 (Netherlands)	338 community adolescents (M/F)	Conflict and Management of Relationships study (CONAMORE; approximately 1 year follow up)	The Children's Depression Inventory (CDI; Kovacs, 1985) (early adolescent group mean age 12.37, and middle adolescence group mean age 16.75)	Self-report delinquency measure (Baerveldt et al., 2003; age not reported)	No	4
Article ID 3	Beyers & Loeber, 2003 (USA)	506 community adolescents (male)	Pittsburg Youth Study (1 year – 4 years)	Short Mood and Feelings Questionnaire (SMFQ; Messer et al., 1995; 13 items summed, report on past 2 weeks); (mean age 13.5)	25 items from Self-Reported Delinquency (SRD; Elliott, Huizinga, & Ageton, 1985) measure past year any offending (measured at ages 14.5, 15.5, 16.5, and 17.5)	No	6

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 4	Blitstein et al., 2005 (USA)	2,335 community adolescents (M/F)	TEENS study (18 month follow up)	CES-D (20 items), scores on 90th percentile cut off for determining depression (beginning of 7th grade)	Self-report measure of violent behavior (Birnbuam et al., 2002; end of 8th grade)	Yes for many individual and contextual variables	6
Article ID 5	Capaldi & Stoolmiller, 1999 (USA)	201 community adolescents (male)	Oregon Youth Study (approximately 4 year follow up)	Child Depression Rating Scale (CDRS; Pozanski, Cook, & Carroll, 1979; assessed at grades 6, 7, 8, and averaged)	Official records of arrests for general offending (grade 12)	No	6
Article ID 6	Caprara et al., 2010 (Italy)	452 community adolescents (M/F)	Italian longitudinal project (approximately 4 year follow up)	CES-D (20 items); (mean age 15.83)	Achenbach Delinquency Scale (Achenbach, 1991b) measuring any offending (mean age 19.80)	No	6
Article ID 7	Cochrane & Viljoen, 2016 (Canada)	152 justice-involved youth (M/F)	Mental Health, Risks, and Strengths Study (1 year follow up)	Personality Assessment Inventory – Adolescent (PAI-A; Morey, 2007) Depression scale (mean age 15.96)	Official records of charges for general and violent offending (mean age 16.96)	No	6
Article ID 8	Copeland et al., 2007 (USA)	1,420 community adolescents (M/F)	Great Smokey Mountain Study (6 – 11 year follow up)	Child and Adolescent Psychiatric Assessment (Angold & Costello, 2000) used algorithm and symptom count to determine diagnosis of depression (assessed between ages 9 – 15)	Official arrest records (age 21)	No	5

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 9	Diamantopoulou et al., 2011 (Netherlands)	1,214 community adolescents (M/F)	Zuid-Holland longitudinal study Waves 1, 3-6 (approximately 14 year follow up)	Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983) parental ratings of childhood depression (age N.R.)	Young Adult Self Report (Achenbach, 1990) any offending (mean age 23.3 years)	No	5
Article ID 10	Elkington et al., 2015 (USA)	1,659 justice-involved adolescents (M/F)	Northwestern Juvenile Project (approximately 2 year follow up)	Diagnostic Interview Schedule for Children Version IV (Shaffer et al., 2000; mean age 18.6.)	Self-reported violence based on the Denver Youth Survey (Institute of Behavioral Science, 1991; mean age 20.2)	No	6
Article ID 11	Herrera, 2001 (USA)	296 high risk adolescents (half justice- involved); (M/F)	Longitudinal study on battered women and their children (follow up approximately at 1.5 violence and 4 years general offending)	CES-D (20 items averaged); (mean age 15.1 years)	Self-report inventory regarding violence (mean age 16.9) and official records of general recidivism (mean age 19 years) any and violent	No	6
Article ID 12	Katsiyannis et al., 2004 (USA)	299 incarcerated adolescents (M/F)	No (3 year follow up)	Reynolds Adolescent Depression Scale (RADS; Reynolds 1987; ages 12.8-18.8, mean age 16.2)	Official records of general offending (age not reported)	No	5

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 13	Mason et al., 2004 (USA)	765 community adolescents (M/F)	Seattle Social Development Project (follow up approximately 11 years)	Teacher's Report Form (Achenbach, 1991a) depression scale (approximately 10 years at baseline)	2 or more violent incidents are reported in the Diagnostic Interview Schedule Version III (Robins, Helzer, Croughan, Williams, Spitzer, 1981) (mean age 21.3 years)	No	5
Article ID 14	Mason et al., 2007 (USA)	429 community adolescents (M/F)	PROJECT Family, waves 1-5; (approximately 1 year follow up)	Child Behavior Checklist (CBCL) youth report (ages 11-15 at 1 year intervals)	Eight items assessing general offending adapted from Self Report of Offending (SRO; Elliott, Huizinga, & Menard, 1989) (ages 12-16 at 1 year intervals)	No	5
Article ID 15	McCarty et al., 2008 (USA)	808 community adolescents (M/F)	Seattle Social Development Project (3-7 year follow up)	Teacher's Report Form (Achenbach, 1991a) depression scale (grade 5)	Self-report measure of general offending (8^{th}, 9^{th}, 10^{th}, and 12^{th} grade)	No	4
Article ID 16	Moon et al., 2009 (South Korea)	787 community adolescents (M/F)	Ongoing longitudinal research (1 year follow up)	Trait based depression scale (Piquero & Sealock, 2000); (grade 8, approximately age 13)	Self-reported violent (6 items) and property (11 items) offending (grade 9, approximately age 14)	Yes (presents additional results after controlled for many individual and contextual variables)	8

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 17	Ostrowsky, 2007 (USA)	947 community adolescents (M/F)	Rochester Youth Development Study waves 2-9 (follow up 3.5 years approximately)	CES-D (14 items averaged); (approximate age 14, wave 2)	Self-reported violent offending in the previous 6 months (collected at waves 3-9, age at wave 9 was 17.5 approximately)	Yes (presents additional results after controlled for alcohol use, past violence, religious ties, school commitment, parental attachment) not able to compare	7
Article ID 18	Overbeek et al., 2006 (Sweden)	126 community adolescents (M/F)	Swedish longitudinal study of adolescence (approximately 2 year follow up)	CES-D (18 items), scores above 75th percentile cut off for determining depression (8th grade)	15 item self-report measure of general offending (10th grade)	No	6
Article ID 19	Pardini et al., 2012 (USA)	503 high risk adolescents (male)	Pittsburg Youth Study (6 year follow up)	Recent Mood and Feelings Questionnaire (13 item self-report measure of past 2 week symptoms using DSM-III-R criteria), data was trichotomized to lower 25%, middle 50%, and upper 25% (age 12)	Self Report of Delinquency (SRD) 4 items capturing violence (measured annually from ages 13-18)	No	5

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 20	Park, 2012 (USA)	898 community adolescents (M/F)	Rochester Youth Development Study waves 7 and 8 (follow up approximately 6 months)	CES-D (14 items averaged); (approximate age 17, wave 7)	Self-reported any offending in the previous 6 months (collected at wave 8, approximate age 17.5)	No	6
Article ID 21	Siennick 2007 (USA)	13,155 community adolescents (M/F)	ADD Health waves 1 and 3 (6 years follow up)	CES-D (9 items) mean of items used (age N.R.)	8 item self-report on any offending (mean age 21.76)	No	6
Article ID 22	Simons et al., 2003 (USA)	718 community adolescents (M/F)	Family and Community Health Study (FACHS); (2 year follow up)	Diagnostic Interview Schedule-IV (Shaffer et al., 2000) used depression symptom counts (ages 10-12)	Diagnostic Interview Schedule-IV – conduct disorder section, used frequency counts of shoplifting, assault, burglary, and robbery (age N.R.)	No	5
Article ID 23	Stuewig & McCloskey, 2005 (USA)	279 high risk adolescents (half justice- involved); (M/F)	Longitudinal study on battered women and their children (follow up approximately at 1.5 self-report and 4 years official records)	Child Assessment Schedule (CAS; Hodges et al., 1981) sum score of symptoms present (mean age 15.1 years)	Official arrest records (mean age 18.6) and self-reported general offending (mean age 16.9 years)	No	6

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 24	Thomas et al., 2017, (USA)	1216 justice-involved adolescents (male)	Crossroads Study of first time offenders (6 month follow up)	The Major Depressive Disorder (MDD) subscale from the Revised Child Anxiety and Depression Scale (Chorpita et al., 2000; mean age 15.3 years)	Self Report of Offending (SRO; age N.R.)	No	6
Article ID 25	Vieno et al., 2008 (Italy)	107 community adolescents (M/F)	Longitudinal Italian study of middle schoolers (10 month follow up)	Italian version of CES-D (17 items) (mean age 12.5)	11 item self-report measure of general offending (Kiesner, 2002; age not reported)	No	5
Article ID 26	Wareham & Boots, 2012 (USA)	1,201 community adolescents (M/F)	Project on Human Development in Chicago Neighbourhoods (PHDCN) waves 1 and 2 (approximately 2 year follow up)	Youth Self Report (YSR) Affective problems scale (Achenbach,1991; age 13)	Self-reported frequency of engaging in six different types of violent offenses (age N.R.)	No	6
Article ID 27	Weaver et al., 2008 (USA)	88 community adolescents (M/F)	Notre Dame Adolescent Parenting Project (IVDAPP; approximately 4 year follow up)	Children's Depression Inventory (CDI) (age 10)	16 item self-report measure of general offending, and Violent Behaviors Scale (UNC Carolina Population Center, 2003) (age 14)	No	5

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 28	You & Lim 2015 (South Korea)	2013 community adolescents (M/F)	Korean Children and Youth Panel Study (approximately 1 year follow up)	Symptom Checklist-90-Revised (Derogatis, 1983), a self-report measure of mental illness. 13 items from depression scale used (age 11 approximately)	Self-reported violent and non-violent offending (age 12 approximately)	No	5
Article ID 29	Yu et al., 2017 ALSPAC sample (UK)	4,030 community adolescents (M/F)	Avon Longitudinal Study of Parents and Children (ALSPAC); (follow up approximately 4 years)	Short Mood and Feelings Questionnaire (SMFQ); (approximately age 13)	Self-report and mother report violent offending questions based on Study of Youth Transitions and Crime. If youth committed 1 violent offence in ages 14-17 categorized as violent.	Yes (presents additional results after controlling for family SES and previous violence)	6
Article ID 29	Yu et al., 2017 FBC sample (Finland)	59,476 adolescent population birth cohort (M/F)	Finnish Birth Cohort 1987 (FBC); (follow up unclear)	At least two outpatient diagnoses of major depression (ICD-10 criteria) (age ranged from 11-25 years, mean age at first diagnosis 19.7)	Official records of violent convictions (age unclear)	Yes (presents additional results after controlling for family SES and previous violence)	7

	Authors, Year (country)	Sample (gender)	Part of a Larger Study? (follow up time)	Measurement of Depression (age at measurement)	Measurement of Offending (age at measurement)	Were other Variables Controlled for?	Newcastle – Ottawa Scale Rating
Article ID 29	Yu et al., 2017 RADAR sample (Netherlands)	678 community adolescents (M/F)	Research on Adolescent Development and Relationships (RADAR); (follow up approximately 4 years)	Reynolds Adolescent Depression Scale-2 (RADS-2; Reynolds, 2002), 23 item self-report measure of depression (mean age 13.1)	Self-report violent offending questions based on International Self Report Delinquency Study. If youth committed 1 violent offence in ages 14-17 categorized as violent.	Yes (presents additional results after controlling for family SES and previous violence)	6

Quality of Included Studies

The mean study quality score of the studies was 5.71 (*SD* = 0.86) out of 8 on the Newcastle-Ottawa Scale, representing moderate to high methodological quality overall. The majority of studies fell in the high quality range (*k* = 19, 61.3%), with the remaining studies (*k* = 12, 38.7%) falling in the moderate range. Most studies (*k* =26, 83.9%) lost points on the comparability item which assesses whether the study controlled for potential confounding variables in the analyses. The mean score was 0.39 (SD = 0.76) out of 2, with 77.4% (*k* = 24) of studies scoring zero. Although many of the included studies controlled for variables such as socioeconomic status or substance use in analyses, if in the analyses of interest (i.e., the relationship between depression and offending), no variables were controlled for, then the studies would be given a zero on the comparability item. Therefore, this item captures the methodological rigor of each study in specifically addressing the present research questions, rather than the study as a whole. A detailed quality assessment is presented in Table 2.

Table 2 Newcastle Ottawa Study Quality Scores

		Selection			Comparability		Outcome		Total Score
		Representative-ness of Exposed Cohort (depressed participants)	Selection of Non-exposed Cohort (non-depressed participants)	Ascertainment of Exposure (quality of depression measurement)	Comparability of Cohorts (did they control for possible confounds?)	Assessment of Outcome (quality of offending measurement)	Was Follow Up Long Enough for Outcomes to Occur?	Adequacy of Follow up Cohorts (is retention rate reasonable)	
Article ID 1	Anderson, 2015 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Yes controlled for many confounds (2 points)	Self-report (1 point)	Yes, 13 year follow up (1 point)	Not reported (0 point)	7 points
Article ID 2	Aske et al., 2007 (Netherlands)	Not representative, selected a subsample of participants based on personality traits (0 points)	Subsample selected based on personality traits 0 points)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 1 year follow up (1 point)	96-99% retention rate (1 point)	4 points
Article ID 3	Beyers & Loeber, 2003 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 1- 4 year follow up (1 point)	83.2% retention rate (1 point)	6 points
Article ID 4	Blitstein et al., 2005 (USA)	Somewhat not representative, large amounts of unusable data (0 point)	Drawn from the same community (1 point)	Self-report (1 point)	Yes controlled for many confounds (2 points)	Self-report (1 point)	Yes, 18 month follow up (1 point)	75% retention, with major differences (0 point)	6 points
Article ID 5	Capaldi & Stoolmiller, 1999 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 1-4 year follow up (1 point)	98% retention rate (1 point)	6 points
Article ID 6	Caprara et al., 2010 (Italy)	Truly representative (1 point)	Drawn from same community (1 point)	Structured Interview (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 4 year follow up (1 point)	Complete retention (1 point)	6 points
Article ID 7	Cochrane & Viljoen, 2016 (Canada)	Truly representative (1 point)	Drawn from same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Official records (1 point)	Yes, 1 year follow up (1 point)	Complete retention (1 point)	6 points
Article ID 8	Copeland et al., (2007)	Truly representative (1 point)	Drawn from same community (1 point)	Structured Interview (1 point)	Did not control for any confounds (0 point)	Official Records (1 point)	Yes, 5 year follow up (1 point)	83% retention, no analyses	5 points

		Representative-ness of Exposed Cohort (depressed participants)	Selection Selection of Non-exposed Cohort (non-depressed participants)	Ascertainment of Exposure (quality of depression measurement)	Comparability Comparability of Cohorts (did they control for possible confounds?)	Assessment of Outcome (quality of offending measurement)	Outcome Was Follow Up Long Enough for Outcomes to Occur?	Adequacy of Follow up Cohorts (is retention rate reasonable)	Total Score
Article ID 9	Diamantopoulou et al., 2011 (Netherlands)	Truly representative (1 point)	Drawn from same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 10 year follow up (1 point)	80% retention, attrition correlated with increased aggression (0 point)	5 points
Article ID 10	Elkington et al., 2015 (USA)	Truly representative (1 point)	Drawn from same community (1 point)	Structured Interview (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 2 year follow up (1 point)	85.3% retention (1 point)	6 points
Article ID 11	Herrera, 2001 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 1.5 – 4 year follow up (1 point)	Complete retention (1 point)	6 points
Article ID 12	Katsiyannis et al., 2004 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Official records of juvenile offending, but adult offences not available (0 point)	Yes, 3 year follow up (1 point)	Complete retention (1 point)	5 points
Article ID 13	Mason et al., 2004 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Teacher report (0 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 11 year follow up (1 point)	95% retention rate (1 point)	5 points
Article ID 14	Mason et al., 2007 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 1 year follow up (1 point)	67-73% retention rate, no analyses (0 point)	5 points
Article ID 15	McCarty et al., 2008 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Teacher report (0 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 3-7 year follow up (1 year)	Not reported (0 point)	4 points
Article ID 16	Moon et al., 2009 (South Korea)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Controlled for many confounds (2 points)	Self-report (1 point)	Yes, 1 year follow up (1 point)	84% retention, no differences (1 point)	8 points

		Representative-ness of Exposed Cohort (depressed participants)	Selection Selection of Non-exposed Cohort (non-depressed participants)	Ascertainment of Exposure (quality of depression measurement)	Comparability Comparability of Cohorts (did they control for possible confounds?)	Assessment of Outcome (quality of offending measurement)	Outcome Was Follow Up Long Enough for Outcomes to Occur?	Adequacy of Follow up Cohorts (is retention rate reasonable)	Total Score
Article ID 17	Ostrowsky, 2007 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Controlled for many confounds (2 points)	Self-report with limited questions (0 point)	Yes, 3.5 year follow up (1 point)	92% retention rate (1 point)	7 points
Article ID 18	Overbeek et al., 2006 (Sweden)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 2 year follow up (1 point)	89% retention rate (1 point)	6 points
Article ID 19	Pardini et al., 2012 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 6 year follow up (1 point)	Not reported (0 point)	5 points
Article ID 20	Park, 2012 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 6 month follow up (1 point)	83% retention (1 point)	6 points
Article ID 21	Siennick 2007 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 6 year follow up (1 point)	92% retention (1 point)	6 point
Article ID 22	Simons et al., 2003 (USA)	Unclear selection process (0 point)	Drawn from the same community (1 point)	Official diagnosis from clinician interview (1 point)	Did not control for any confounds (0 point)	Clinician interview (1 point)	Yes, 2 year follow up (1 point)	86% retention (1 point)	5 points
Article ID 23	Stuewig & McCloskey, 2005 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Official diagnosis (1 point)	Did not control for any confounds (0 point)	Official records and self-report (1 point)	Yes, 1.5 year follow up with self-report, and 4 year follow up official records (1 point)	Complete retention (1 point)	6 points
Article ID 24	Thomas et al., 2017, (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for confounds (0)	Self-report (1 point)	Yes, 6 month follow up (1 point)	96% retention (1point)	6 points

		Representative-ness of Exposed Cohort (depressed participants)	Selection Selection of Non-exposed Cohort (non-depressed participants)	Ascertainment of Exposure (quality of depression measurement)	Comparability Comparability of Cohorts (did they control for possible confounds?)	Assessment of Outcome (quality of offending measurement)	Outcome Was Follow Up Long Enough for Outcomes to Occur?	Adequacy of Follow up Cohorts (is retention rate reasonable)	Total Score
Article ID 25	Vieno et al., 2008 (Italy)	Somewhat not representative, large amounts of unusable data (0 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 10 month follow up (1 point)	71% retention rate, conducted analyses (1 point)	5 points
Article ID 26	Wareham & Boots, 2012 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 2 year follow up (1 point)	86% retention (1 point)	6 points
Article ID 27	Weaver et al., 2008 (USA)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 4 year follow up (1 point)	Not reported (0 point)	5 points
Article ID 28	You & Lim 2015 (South Korea)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Did not control for any confounds (0 point)	Self-report (1 point)	Yes, 1 year follow up (1 point)	Not reported (0 point)	5 points
Article ID 29	Yu et al., 2017 ALSPAC sample (UK)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Controlled for many confounds (2 points)	Binary self-report (0 point)	Yes, 4 year follow up (1 point)	Not reported (0 point)	6 points
Article ID 29	Yu et al., 2017 FBC sample (Finland)	Truly representative (1 point)	Drawn from the same community (1 point)	Clinician diagnosis (1 point)	Controlled for SES and previous violence (1 point)	Official records (1 point)	Follow up length unclear (0 point)	Complete retention (1 point)	7 points
Article ID 29	Yu et al., 2017 RADAR sample (Netherlands)	Truly representative (1 point)	Drawn from the same community (1 point)	Self-report (1 point)	Controlled for SES and previous violence (1 point)	Self-report (1 point)	Yes, 4 year follow up (1 point)	Not reported (0 point)	6 points

Assessment of Publication Bias and Influential Data Points

Before discussing overall effects, it is good practice to check the distribution results for possible publication bias (Borenstein et al., 2009). Inspection of funnel plots is a way to assess for publication bias in meta-analysis. This method assumes that studies will be normally distributed around the true mean effect size (Borenstein et al., 2009). If the included studies suffer from publication bias, an asymmetric distribution around the mean will be observed in the plot. Regarding the studies reporting outcomes of any offending, (see Figure 2), a visual inspection of the funnel plot shows some asymmetry indicating some studies appear to be missing in the lower left corner of the plot. This suggests that some smaller studies with negative effect sizes may not have been published. Egger and Smith's (1997) regression test for funnel plot asymmetry also confirms ($Z = 2.47$, $p = .024$)[2] this finding. In order to assess the sensitivity of the results to publication bias, Duval and Tweedie's (2000) trim-and-fill procedure removes and inputs studies to create a symmetrical distribution and re-calculates the weighted effect size to compare against the uncorrected effect size. However, using the trim-and-fill procedure no missing studies were identified, suggesting no evidence of publication bias. It is not uncommon for differing measures of publication bias to yield inconsistent results. In cases like this it is important to synthesize findings across different methods (Borenstein et al., 2009). Accordingly, there is some evidence that smaller studies were missing, however random-effects models award relatively more weight to smaller studies than fixed-effects models, making them less susceptible to publication bias. Therefore, although publication bias may be present, the impact appears to be low. Regarding the studies reporting outcomes of violent offending, (see Figure 3) both visual inspection of the funnel plot and Egger and Smith's (1997) regression test ($Z = 0.72$, $p = .482$) do not suggest any publication bias.

[2] Egger regression uses an alpha level of .10.

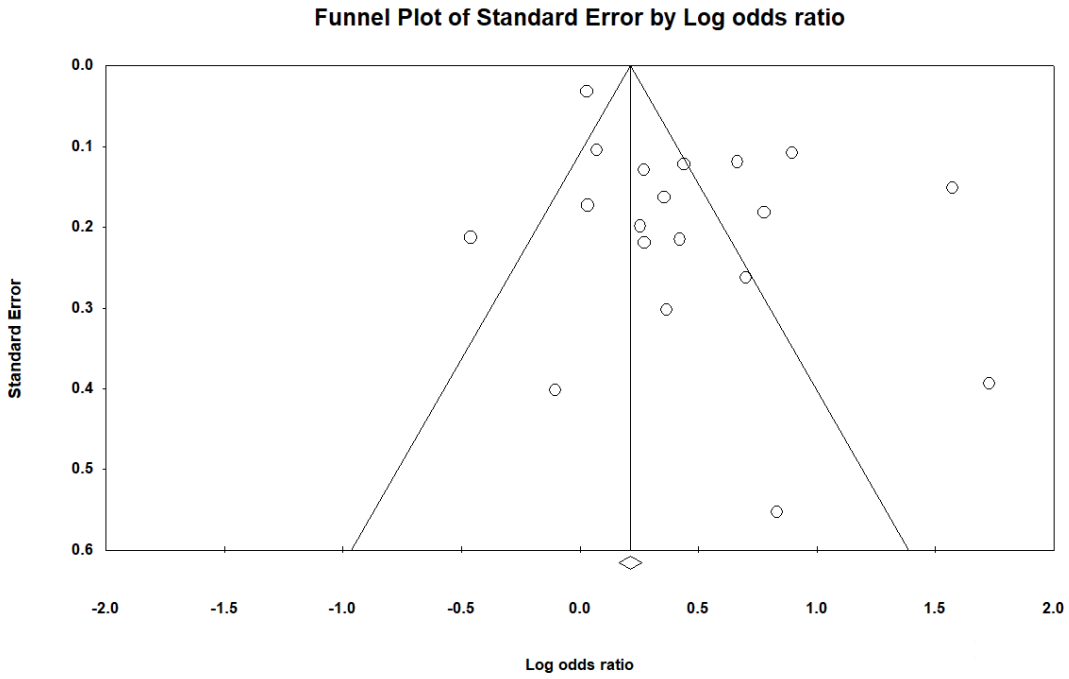

Figure 2 Funnel Plot: Log Odds of Any Offending

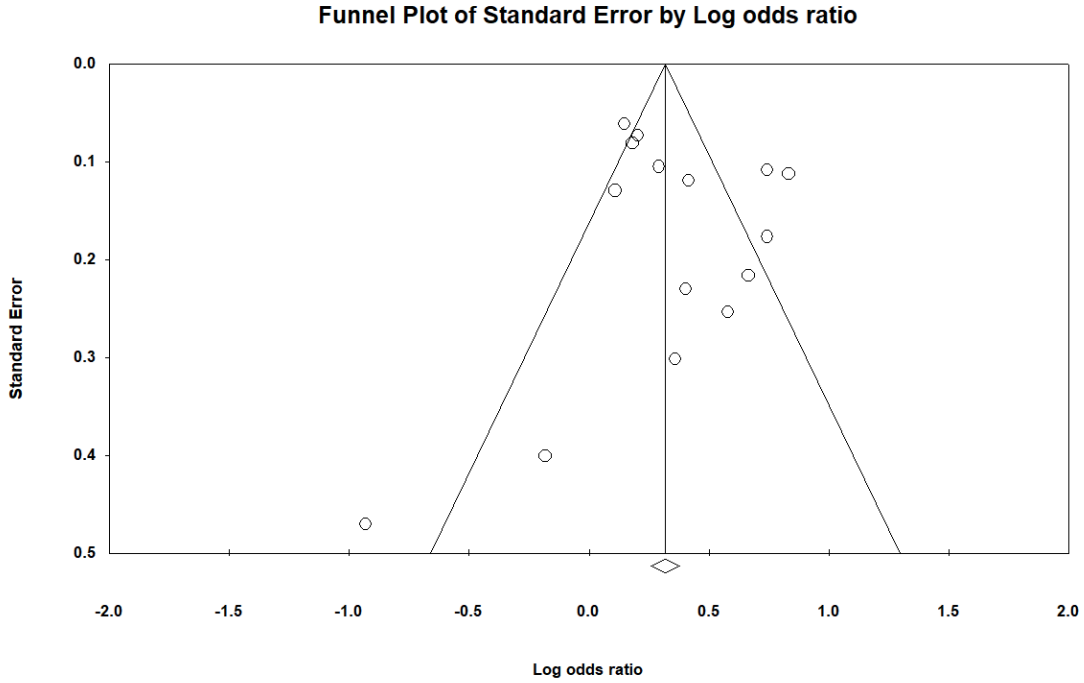

Figure 3 Funnel Plot: Log Odds of Violent Offending

Regarding outliers, visual inspection of the forest plots (see Appendix C), suggested that Simons et al. (2003), and Vieno et al. (2008) were outliers. However, there was little change to the weighted mean effects and no change to the significance of the findings when the studies were removed. Although most researchers agree it is helpful to assess for potential outliers, Hunter and Schmidt (2004) recommend against removing outliers from analyses due to it being difficult to distinguish between true outliers and large sampling error. Therefore, considering both the findings that the outlying data points were not influential and the recommendations against deletion of outliers by experts (e.g., Hunter & Schmidt, 2004; Viechtbauer & Cheung, 2010) consequently the studies were retained in the analyses.

Question 1: What is the Overall Relationship between Depression and Offending?

Any offending. Nineteen study samples examined the relationship between depression and general/any offending. Effect sizes and 95% confidence intervals are presented in Table 3. Wherever possible, studies with different types of offending measurement (i.e., self-report, official records) were analyzed in separate models in order to combine studies with similar outcomes to create a more precise estimate. When combining studies with different types of offending measurement (i.e., self-report, official records, or mixed), the weighted odds ratio was 1.58 (p <.001, 95% CI [1.27, 1.97]), indicating a small positive effect size, such that depression increases the odds of engaging in any offending by 58% (Chen, Cohen, & Chen, 2010). When only studies that used self-report were aggregated (k = 15), the weighted odds ratio was 1.66 (p <.001, 95% CI [1.29, 2.12]). Further, when studies that used official records were aggregated (k = 5), the weighted odds ratio was 1.35 (p = .195, 95% CI [.86, 2.13]). Heterogeneity in all three models was high with significant Q values and I^2 values ranging from 74.97 – 93.09.

Violent offending. Fifteen samples examined the relationship between depression and violent offending. When combining studies with different types of offending measurement (i.e., self-report, official records, or mixed), the weighted odds ratio was 1.45 (p = <.001, 95% CI [1.25, 1.69]), indicating a small positive effect size, such that depression increases the odds of engaging in violent offending by 45%. Of the 15 samples that examined violence, only 1 study used official records, with the other 14 all using self-report to assess violence and therefore,

they were not able to be examined separately. Heterogeneity was high (Q = 70.52, p<.001, I^2 = 80.15).

Table 3 Overall Analyses

Outcomes	*k*	Random-Effect Models O.R.$_w$	95% CI		*Z*	*p*	Heterogeneity *Q*	*p*	I^2
Any Offending									
All methods[1]	19	1.58	1.27	1.97	4.07	<.001	217.44	<.001	91.72
Self-report[2]	15	1.66	1.29	2.12	4.00	<.001	202.59	<.001	93.09
Official records[3]	5	1.35	0.86	2.13	1.30	.195	15.98	.003	74.97
Violent Offending									
All methods[4]	15	1.45	1.25	1.69	4.82	<.001	70.52	<.001	80.15

k = number of effect sizes that were aggregated.

1: Any offending assessed by all methods: Aske et al. (2007); Beyers and Loeber (2005); Capaldi and Stoolmiller (1999); Caprara et al. (2010); Cochrane and Viljoen (2016); Copeland et al. (2007); Diamantopoulou et al. (2011); Herrera (2011); Katsiyannis et al. (2004); Mason et al. (2007); McCarty et al. (2008); Overbeek et al. (2006); Park (2012); Siennick (2007); Simons et al. (2003); Stuewig and McCloskey (2005); Thomas et al. (2017); Vieno et al. (2008); Weaver et al. (2008). 2: Any offending assessed by self-report: Aske et al. (2007); Beyers and Loeber (2005); Caprara et al. (2010); Cochrane and Viljoen (2016); Diamantopoulou et al. (2011); Mason et al. (2007); McCarty et al. (2008); Overbeek et al. (2006); Park (2012); Siennick (2007); Simons et al. (2003); Stuewig and McCloskey (2005); Thomas et al. (2017); Vieno et al. (2008); Weaver et al. (2008). 3: Any offending assessed by official records: Capaldi and Stoolmiller (1999); Copeland et al. (2007); Herrera (2011); Katsiyannis et al. (2004); Stuewig and McCloskey (2005). 4: Violent offending assessed by all methods: Anderson (2015); Blitstein et al. (2005); Cochrane and Viljoen (2016); Elkington et al. (2015); Herrera et al. (2001); Mason et al. (2004); Moon et al. (2009); Ostrowsky (2007); Pardini et al. (2012); Wareham & Boots (2012); Weaver et al. (2008); You and Lim (2015); Yu et al. (2017) ALSPAC, FBC, and RADAR samples.

Question 2: Does the Relationship Vary between Community and Justice-Involved Adolescents?

Studies were dichotomized into two types of settings: community samples (e.g., school students or population samples), and justice-involved samples (e.g., incarcerated, probation, or history of arrests). Analyses for any offending and violent offending were re-run grouping for population type, and subgroup analyses were conducted to determine if any of the differences were significant. Outcomes were combined across all types of measurement (i.e., self-report, official records), due to low numbers of studies in each group.

Any offending. When calculating aggregated odds ratios for any offending, 19 of the studies examined the outcome of any offending, of which 14 included samples of community adolescents, and 5 studies included samples of justice-involved adolescents (see Table 4). The weighted odds ratio for community samples was 1.66 (p <.001, 95% CI [1.29, 2.14]), and 1.37 for justice-involved samples (p =.233, 95% CI [.82, 2.28]), however the difference between the two populations did not reach significance. This indicates that in community samples, depression is a small but significant risk factor for any offending, but in justice-involved samples, the relationship between depression and any offending did not reach statistical significance.

Violent offending. When calculating weighted odds ratios for violent offending, 12 studies included samples of community adolescents, however only 2 studies examined violent offending in justice-involved samples, which therefore does not permit an aggregated effect size to be calculated for only justice-involved samples. The weighted odds ratio for community samples was 1.42 (p = <.001, 95% CI [1.20, 1.68]), indicating that depression is a small but significant risk factor for violent offending in community sample

Table 4 Analyses by Population

Outcomes	Random-Effect Models						Heterogeneity		
	k	O.R.$_w$	95% CI		*Z*	*p*	*Q*	*p*	I^2
Community Samples									
Any Offending[1a]	14	1.66[a]	1.29	2.14	3.92	<.001	164.22	<.001	92.08
Violent Offending[2b]	12	1.42	1.20	1.68	4.08	<.001	67.70	<.001	83.75
Justice-Involved Samples									
Any Offending[3a]	5	1.37[a]	0.82	2.28	1.19	.233	35.18	<.001	86.63

Note. *k* = number of effect sizes that were aggregated

1: Community sample, any offending assessed by all methods: Aske et al. (2007); Beyers and Loeber (2005); Capaldi and Stoolmiller (1999); Caprara et al. (2010); Copeland et al. (2007); Diamantopoulou et al. (2011); Mason et al. (2007); McCarty et al. (2008); Overbeek et al. (2006); Park (2012); Siennick (2007); Simons et al. (2003); Vieno et al. (2008); Weaver et al. (2008). 2: Community sample, violent offending assessed by all methods: Anderson (2015); Blitstein et al. (2005); Cochrane and Viljoen (2016); Mason et al. (2004); Moon et al. (2009); Ostrowsky (2007); Pardini et al. (2012); Wareham & Boots (2012); Weaver et al. (2008); You and Lim (2015); Yu et al. (2017) ALSPAC, FBC, and RADAR samples. 3: Justice-involved sample, any offending assessed by all methods: Cochrane and Viljoen (2016); Herrera (2011); Katsiyannia et al. (2004); Stuewig and McCloskey (2005); Thomas et al. (2017).

a: Difference between community and justice-involved samples was not significant ($Q(1) = 1.97$, $p = .160$).

Question 3: Does the Relationship Vary between Males and Females?

Any offending. When studies were examined reported effect sizes separately for each sex, six studies for males and five studies for females were available (see Table 5). When combining studies that used any method to assess offending, the weighted odds ratio was 1.71 for males (p = .012, 95% CI [1.27, 2.61]), and 1.74 for females (p = .158, 95% CI [0.81, 3.75]). The difference between males and females was non-significant ($Q(1) = 0.01$, p = .974). This indicates that depression is a small but significant risk factor for general offending among males. Although the effect size was slightly larger for females, it did not reach statistical significance. This may be due to a 25% smaller aggregated sample size for females (e.g., n = 1266 for females as compared to n = 1707 for males), and higher levels of heterogeneity across the effect sizes.

Violent offending. Regarding violent offending, eight studies for males and seven studies for females were available (see Table 5). When combining studies that used any method to assess violence, the weighted odds ratio was 1.91 for males (p <.001, 95% CI [1.44, 2.55]), and 2.40 for females (p < .001, 95% CI [1.53, 3.79]). The difference between males and females was non-significant ($Q(1) = 6.89$, p =.406). This indicates that depression is a small, but significant risk factor for violent offending in both males and females.

Table 5 Analyses by Sex

Outcomes	*k*	O.R.$_w$	Random-Effect Models 95% CI		Z	*p*	Heterogeneity *Q*	*p*	I^2
Males Only									
Any Offending[1]	6	1.71[a]	1.27	2.61	2.52	.012	24.81	<.001	79.85
Violent Offending[2]	6	1.91[b]	1.44	2.55	4.43	<.001	14.97	.010	66.60
Females Only									
Any Offending[3]	5	1.74[a]	0.81	3.75	1.41	.158	48.74	<.001	91.79
Violent Offending[4]	5	2.40[b]	1.53	3.79	3.78	<.001	18.32	.001	78.16

Note. *k* = number of effect sizes that were aggregated

1: Males, any offending assessed by all methods: Beyers and Loeber (2005); Caprara et al. (2010); Mason et al. (2007); McCarty et al. (2008); Simons et al. (2003); Weaver et al. (2008). 2: Males, violent offending assessed all methods: Ostrowsky (2007); Simons et al. (2003); Weaver et al. (2008); Yu et al. (2017) ALSPAC, FBC, and RADAR samples. 3: Females, any offending assessed by all methods: Caprara et al. (2010); Mason et al. (2007); McCarty et al. (2008); Simons et al. (2003); Weaver et al. (2008). 4: Females, violent offending assessed by all methods: Ostrowsky (2007); Weaver et al. (2008); Yu et al. (2017) ALSPAC, FBC, and RADAR samples. a: Difference between males and females not significant ($Q(1) = 0.01$, $p = .974$). b: Difference between males and females not significant ($Q(1) = 6.89$, $p = .406$).

Question 4: Which Factors Moderate the Effect of Depression on Offending?

Given the high heterogeneity across almost all analyses, moderator variables were analyzed to determine whether any study variables may be helpful in accounting for the high variability between studies.

Follow up Time. Study follow up time was highly variable and ranged from 6 months to 21 years. Only one study reported a mean follow up time across participants; the remaining studies reported approximate follow up times. Therefore, this did not allow follow up time to be analyzed as a continuous variable. Accordingly, follow up time was trichotomized into short (2 years or less), mid (2 years 1 month – 5 years), and long (greater than 5 years) follow up. Six studies were not able to be categorized given the follow up time varied significantly between participants (e.g., follow up ranged from 3 months – 5 years). Outcomes were collapsed over the type of measurement of offending (e.g., self-report or official records) due to low numbers of studies (see Table 6). For any offending, the weighted odds ratio for was 2.42 for short follow up (p =.001, 95% CI [1.46, 4.00]) and 1.04 for long follow up lengths (p =.355, 95% CI [.96, 1.14]). The difference between short and long follow up lengths was significant ($Q(1) = 10.46$, p =.001). This indicates that in a short follow up (2 years or less) depression is a significant risk factor for any offending; however, the association is no longer significant after the follow up length is increased (beyond 2 years).

For violent offending, a similar pattern occurred. The weighted odds ratio for violent offending was 1.19 for short follow up (p <.001, 95% CI [1.10, 1.30]), 1.83 for mid length follow up (p <.001, 95% CI [1.29, 2.60]), and 1.10 over long follow ups (p =.710, 95% CI [.64, 1.91]). The difference between short, mid, and long follow up was not significant ($Q(2) = 5.62$, p =.060). The Q statistics and I^2 values indicated large amounts of heterogeneity in all the models.

Table 5 Analyses by Follow up Time

		Random-Effect Models					Heterogeneity		
Outcomes	***k***	**O.R.$_w$**	**95% CI**		**Z**	***p***	***Q***	***p***	**I^2**
Short Follow Up									
Any Offending[1]	5	2.42[a]	1.46	4.00	3.45	.001	48.83	<.001	91.81
Violent Offending[2]	4	1.19[b]	1.10	1.30	4.19	<.001	1.67	.643	0.00
Mid Follow Up									
Violent Offending[4]	3	1.83[b]	1.29	2.60	3.40	.001	5.04	.081	60.30
Long Follow Up									
Any Offending[5]	3	1.04[a]	0.96	1.14	0.93	.355	2.25	.325	10.91
Violent Offending[6]	3	1.10[b]	0.64	1.91	0.37	.710	7.99	.018	74.97

Note. *k* = number of effect sizes that were aggregated

1: Short follow up for any offending outcome: Aske et al. (2007); Overbeek et al. (2006); Park (2012); Simons et al. (2003); Vieno et al. (2008). 2: Short follow up for violent offending outcome: Blitstein et al. (2005); Moon et al. (2009); Wareham & Boots (2012); You and Lim (2015). 3: Mid follow up period for any offending outcome: Caprara et al. (2010); Weaver et al. (2008). 4: Mid follow up period for violent offending outcome: Weaver et al. (2008); Yu et al. (2017) ALSPAC and RADAR samples. 5: Long follow up period for any offending outcome: Copeland et al. (2007); Diamantopoulou et al. (2011); Siennick (2007). 6: Long follow up period for violent offending outcome: Anderson (2015); Mason et al. (2004); Pardini et al. (2012).

a: Difference between short and long length follow up was significant ($Q(1) = 10.46$, $p = .001$). b: Difference between short, mid, and long follow up was not significant ($Q(2) = 5.62$, $p = .060$).

Controlling for Confounds. A small number of studies reported effect sizes before and after a number of confounding variables were controlled for (see Table 7). Moon et al. (2009), and Yu et al. (2017; ALSPAC, RADAR, and FBC samples) were aggregated to compare the effect size for violent offending before and after controlling for variables. Not enough studies reported this for another outcome to allow comparison. Moon et al. (2009) presented effect sizes before and after controlling for gender, academic achievement, and previous offending. Yu and colleagues (2017) reported effect sizes for all three samples before and after controlling for previous violence and family socioeconomic status. The weighted odds ratio for violent offending across the four samples was 1.83 before controlling (p <.001, 95% CI [1.32, 2.56]), and 1.56 after controlling (p =.020, 95% CI [1.07, 2.28]) for confounds. Although the odds ratio was smaller after controlling for potential confounds, the difference between the two effect sizes was not significantly different.

Table 6 Analyses Before and After Controlling for Confounds

		Random-Effect Models					Heterogeneity		
Outcomes	***k***	**O.R.$_w$**	**95% CI**		**Z**	***p***	***Q***	***p***	**I^2**
Before Controlling[1]									
Violent Offending	4	1.83[a]	1.32	2.56	3.59	<.001	20.86	<.001	85.61
After Controlling[1]									
Violent Offending	4	1.56[a]	1.07	2.28	2.32	.020	23.75	.001	87.37

Note. *k* = number of effect sizes that were aggregated.

1: Studies that report effect sizes before and after controlling for confounds: Moon et al. (2009), Yu et al. (2017) ALSPAC, FBC, and RADAR samples.

a: Difference between before and after controlling for confounds was not significant ($Q(1) = 0.38$, $p = .538$).

Study Quality and Other Publication Variables. In order to examine moderator variables that are continuous, meta-regressions were employed. Meta-regression requires a minimum of 10 studies per moderator variable (Borenstein et al., 2009); therefore, with 29 studies in total, a maximum of 3 moderator variables could be analyzed. The total score from the Newcastle-Ottawa study quality was analyzed as well as the year of publication and the country of publication (dichotomized as USA/non-USA samples; see Table 8). No variables were significant, indicating that none moderated the outcome between depression and general or violent offending.

Table 7 Meta-regression of Study Quality and Other Variables

	Random-Effect Models					
Outcomes	***β***	**S.E.**	**95% CI**		**Z**	***p***
Any Offending[1]						
Publication Year	-0.07	0.05	-0.16	0.03	-1.33	.182
Country: USA	-0.08	0.31	-0.69	0.53	-0.27	.778
Newcastle Ottawa Score	-0.01	0.15	-0.30	0.29	-0.04	.972
Model heterogeneity statistics	k = 14, Q = 126.59, I^2 = 92.10, p<.001, R^2 = 0.00					
Violent Offending[2]						
Publication Year	0.06	0.03	-0.02	0.11	1.54	.069
Country: USA	-0.00	0.22	-0.43	0.43	-0.00	.999
Newcastle Ottawa Score	0.09	0.08	-0.07	0.25	1.09	.277
Model heterogeneity statistics	k = 12, Q = 8.19, I^2 = 79.41, p=.042, R^2 = 0.00					

Note. k = number of effect sizes that were aggregated

1: Any offending assessed by all methods: Aske et al. (2007); Beyers and Loeber (2005); Capaldi and Stoolmiller (1999); Caprara et al. (2010); Diamantopoulou et al. (2011);); Mason et al. (2007); McCarty et al. (2008); Overbeek et al. (2006); Park (2012); Siennick (2007); Simons et al. (2003); Vieno et al. (2008); Weaver et al. (2008). 2: Violent offending assessed by all methods: Anderson (2015); Blitstein et al. (2005); Mason et al. (2004); Moon et al. (2009); Ostrowsky (2007); Pardini et al. (2012); Weaver et al. (2008); You and Lim (2015); Yu et al. (2017) ALSPAC, FBC, and RADAR sample

Chapter 4.

Discussion

To synthesize the literature, a meta-analysis was conducted to examine the relationship between depression and offending in adolescents. Thirteen databases were systematically searched, and reference lists were reviewed. The review captured 29 studies with 97,316 participants from 27 non-overlapping samples. The majority of the studies sampled community populations broadly examining mental health and risk taking in adolescent school students, with a smaller minority examining justice-involved youth, or outpatient mental health patients.

Key Findings

Overall, the current study found that depression is significantly positively associated with any offending, violent offending, and intimate partner violence. However, given the impact of depression varies by sample, these results are best interpreted in context of the differences between community and justice-involved adolescents.

Depression *is* Related to Offending in Community Samples. The findings of the meta-analysis provide evidence that depression is associated with a 66% increased likelihood of general offending and a 42% increased likelihood of violent offending in community samples of adolescents. This finding is in line with what the acting-out model would predict, which reasons that depression causes offending by an individual "acting out" their depressive feeling such as by engaging in risk taking or offending (Wolff & Ollendick, 2006). Specifically, the symptom of irritable mood has been hypothesized to drive this relationship. Wolff and Ollendick (2006) suggest that adolescents with depression may develop conduct problems as the irritability associated with their depression increases in severity. The individual has a more difficult time regulating their irritable mood, which increases conflict with others, rule-breaking, and behaviours such as aggression. Stringaris, Maughan, Copeland, Costello, and Angold (2013) examined the role that irritable mood in the context of depression plays in the development of conduct problems in adolescents. A mixed irritable and depressed mood characterized 41% of the depressed adolescent's affect, and adolescents with mixed

irritable/depressed mood had an 85% increased likelihood of developing conduct problems compared to adolescents with only depressed mood. Although conduct problems are not equivalent to offending, this study provides some tentative evidence that irritability may play an important role in the relationship between depression and offending, and should be further investigated in samples of justice-involved youth.

Given the evidence supports that depression is related to an increased likelihood of offending in community adolescent samples, early identification should be a priority in order to intervene before negative consequences like offending occur. Universal depression screening for children and adolescents is somewhat controversial due to limited low-quality evidence that supports improvements in health outcomes. The Canadian Task Force on Preventative Care recommends against screening due to the lack of evidence; however, the Canadian Paediatric Society endorses routine screening during general health visits for adolescents (Roseman et al., 2017). Further, a 2017 systematic review could not identify a single randomized control trial regarding whether depression screening improves depression outcomes in children and adolescents (Roseman et al., 2017). In light of the lack of evidence, and concerns regarding potential harms of universal screening such as overdiagnosis and overtreatment in the context of healthcare systems that struggle to provide care to people with known diagnoses, the current conclusion is that universal screening should not be implemented. A further barrier to implementing universal screening is a lack of evidence regarding the accuracy of screening tools for depression in adolescents. Roseman and colleagues (2016) conducted a systematic review and found that the most popular screening tools, the Beck Depression Inventory- Primary Care Version (BDI-PC: Steer, Cavalieri, Leonard, & Beck, 1999), and the Patient Health Questionnaire for Adolescents (PHQ-A: Kroenke, Spitzer, & Williams, 1999) had a dearth of research that supported screening accuracy, with most studies not even reporting cut off scores used. In sum, the current state of the literature suggests that further research needs to be conducted before any conclusions can be drawn about how to screen for depression.

Instead, options to improve the early identification of depression that are recommended include training healthcare professionals to recognize, assess, and treat depression, and improving access to mental health services. A promising method to prevent the onset of depression is group-based Cognitive Behavioural Therapy (CBT). An 8-week course of group-based CBT for adolescents at a higher risk for depression

has been shown to be effective in reducing the risk of developing depression from 33% with care as usual, to 21% within the next year (Garber et al. 2009).

Depression *Does Not* Predict Reoffending in Youth Who are Already Involved in the Justice System. When justice-involved samples were compared to community samples in the present meta-analysis, differences emerged. In community samples, depression was associated with a 64% increased likelihood of general offending, compared to a non-significant 37% increased likelihood of re-offending in justice-involved samples. These results are consistent with Bonta and colleagues' (1998; 2014) findings, which showed no significant relationship between mood disorders and offending in adult offenders with mental disorders. Bonta et al. (1998) reported a *Zr* of -.04 regarding the relationship between mood disorders and general recidivism and in the 2014 follow up study, Bonta et al., reported a *d* of -.16 between mood disorders and general recidivism, finding both times that mood disorders and offending are not significantly related.

These results may be explained by the major differences between justice-involved and community samples. The community based samples included in the current review typically assessed school-age children, starting early in adolescence and up until early adulthood. Although the studies did not assess for or control for previous offending, it is likely that the majority of offending captured in these studies represents an initial onset of offending. In contrast, the justice-involved samples captured in the current review assessed adolescents who re-offended. Given initial onset of offending typically occurs in earlier adolescence (Fagan & Western, 2005), it is difficult to tease apart the role that developmental stage and initial onset (compared to maintenance) contribute (Jolliffe, Farrington, Piquero, MacLeod, & van de Weijer, 2017). It has been well established that factors associated with the first incidence of a negative outcome may differ between factors that predict subsequent occurrences (Rutter & Sroufe, 2000). This may provide evidence that depression is a meaningful factor in predicting the initial onset of offending in adolescents, but not predictive of offending in adolescents who have already offended.

There are several explanations that may account for the lack of meaningful relationship between depression and offending in justice-involved adolescents. First, in the current review, only six studies sampled justice-involved participants, and there was

significant heterogeneity among studies. Out of the six studies, five found depression to be a risk factor for general offending, with odds ratios ranging from 1.23 to 2.30. The sixth study, Katsiyannis et al. (2004), which studied incarcerated adolescent males, found depression to be protective against general offending with an odds ratio of 0.63. Due to the small number of studies, future research should examine this further to (a) determine whether the non-significant result is replicated with more studies, and (b) to investigate potential causes of the large variability among the findings.

Depression may, however, be simply a less important factor in predicting offending in justice-involved youth due to increased complexity and needs of that population. It is well documented that youth with justice contact are more likely to have mental disorders, a history of maltreatment, and substance use problems than their same-aged peers (Duran-Bonavila, Vigil-Colet, Cosi, & Morales-Vives, 2017; Kenny, Nelson, & Lennings, 2007). The more risk factors an adolescent has, the less any one individual risk factor may play a role in contributing to offending. The dominant model of offender rehabilitation, the RNR model, currently considers depression as a responsivity factor, meaning that it has no causal relationship to criminal behaviour, but could be considered as part of a rehabilitation plan if depression interfered with treatment of other well-established risk factors, like substance use (Andrews & Bonta, 2010). In this way, depression is a side consideration rather than a focal point. The weight that depression is given must be considered in comparison to other risk factors. A recent meta-analysis synthesized the effect sizes of risk factors for offending in justice-involved youth allowing for comparison. Scott and Brown (2018) found that the effect sizes for males, on variables like poor parenting (d = .21), antisocial personality (d = .37), antisocial peers (d = .32), and education problems (d = .52) were in the small to moderate range and all significantly predicted general recidivism. For females, Scott and Brown (2018) found effect sizes ranging from small to moderate as well for variables such as poor parenting (d = .29), antisocial personality (d = .42), anti-social peers (d = .27), and education problems (d = .52).In the current study, depression achieved a d of .17 in predicting general offending in justice-involved youth. It appears that the effect size estimate for depression in the present meta-analysis is smaller than numerous risk factors considered to be well established. This is consistent with the RNR model's stance that depression is best considered as a non-criminogenic need.

Although depression may not be highly related to recidivism in justice-involved adolescents, this does not mean that it is not important to address. Depression has been found to be highly predictive of other negative outcomes such as self-harm and suicide in these populations (Ahrens & Rexford, 2002, Callaghan et al., 2003, Carswell et al., 2004). Of particular importance are the high prevalence rates of self-harm, suicide ideation and suicide attempts in justice-involved adolescents. Among a sample of adolescents recently assigned to probation, 11.2% self-reported a previous suicide attempt (Langhinrichsen-Rohling & Lamis, 2008), and another study of incarcerated adolescents found that 34% had experienced previous suicide ideation (Ruchkin, Schwab-Stone, Koposov, Vermeiren, & King, 2003). One of the main mandates of the Youth Criminal Justice Act (YCJA) in Canada is to rehabilitate youth to allow them to re-integrate back into society, which is primarily accomplished by targeting criminogenic needs (Department of Justice, 2017). However, every professional working with youth in this setting also has the duty to prevent suicide (White, 2014). Despite this duty, there is evidence that adolescent offenders do not receive referrals for interventions for depression, self-harm, or suicide ideation (Callaghan et al., 2003; Carswell et al., 2004; Chitsabasem et al., 2006). For example, Chistabasem et al. (2006) examined whether young offenders, both on probation and incarcerated, had their criminogenic and mental health needs met. Of the adolescents who were assessed as having a mental health need for treatment of depression 75% did not receive any treatment, and similarly 71% of adolescents with a need for treatment of self-harm did not receive any. This indicates a strong need for more attention to treating depression and related issues such as self-harm and suicide in justice-involved youth.

Depression is Equally Important for Both Sexes. The current study found that depression is of similar importance for predicting both general and violent offending in both males and females; differences were not significant. In males, depression was associated with a 71% increased likelihood of general offending, and in females a 74% increased likelihood of general offending. Regarding violent offending, depression was associated with an 91% increased likelihood in males, and an 140% increased likelihood for females.

Some previous studies that have compared the role of depression and offending between the sexes has found that depression is a weaker and more inconsistent risk factor for females (e.g., Kim & Kim, 2005) compared to males. One explanation of

previous findings is the gender-specific risk enhancement theory, which argues that in phenomenon with an unequal sex ratio, like depression, the demographic with the lower prevalence (boys in the case of depression) are more severely affected by the disorder (Loeber & Keenan, 1994). However, regarding the findings of the current review, the gender-specific risk enhancement theory is not supported, due to the contrary findings that depression was slightly more predictive for females.

Another reason why past research may have found depression to be a stronger risk factor in males could be simply due to base rates of offending. Males tend to engage in higher rates of criminal behaviour compared to females; thus, when measuring the effect of a risk factor on offending, it may be more difficult to detect risk factors for females (Loeber & Keenan, 1994). However, as the current study found, on every single model, females had comparable or slightly higher odds ratios than males. Yet when examining any offending, despite females matching the odds ratios of males, the models were not significant. This is due to the general disparity in sample sizes when assessing males and females, as well as lower base rates of offending in females. Specifically, in the models predicting any offending, females had a *n* of 1416, whereas males had about 25% more with an *n* of 1879. Further, for example, one of the studies included in these models, Herrera (2002), reported that the males had a higher base rate for offending, and that the males were significantly more likely to engage in every type of offending they assessed. Therefore, it is possible that past studies have concluded that depression is more strongly linked to offending in males, simply due to having higher power for analyses for males than with females.

Other researchers have highlighted the fact that females involved in the justice system have higher rates of mental illness, substance use, and trauma history than males (Kerig & Schindler, 2013). Due to these differences, gender-responsive theories reason that mental health issues such as depression might affect females differently than males and differentially lead to offending for girls through increases in risk taking or substance use (Kerig & Becker, 2012). The current study supports depression being predictive for both males and females, however, it is yet to be determined whether the mechanisms that drive the association between depression and offending are the same for males and females.

Limitations and Strengths

As always, a meta-analysis is limited by the quality of research that it draws from. The literature regarding community-based adolescents was sufficiently large enough to draw some general conclusions regarding how depression may be a risk factor for offending. In contrast, the literature regarding individuals who are already involved in the justice system was quite limited, with further research needed to be conducted to determine any consensus. Although some questions were able to be evaluated, such as the differences between community based and justice-involved adolescents, and potential sex differences, some potential moderating variables were not able to be tested. For instance, the difference between self-reported or official diagnoses of depression was not able to be examined due to only one sample (Yu et al., FBC sample, 2017) using official diagnoses to measure depression. Further, only four studies in the current review presented effect sizes before and after controlling for potential confounding variables, thus limiting the extent to which the shared risk factor model could be tested.

Regarding strengths, this is the first meta-analytic review regarding offending, that the author is aware of, that examines depression in isolation, as compared to grouped with other mood disorders. Further, broad inclusion criteria were used in order to capture all types of offending, and all types of populations who may offend, thus increasing the generalizability of the findings. Lastly, the results of the current review provide the first amalgamation of the literature on the link between depression and offending among adolescents, thus providing many new future research directions.

Future Directions

Due to the large unexplained heterogeneity of results between studies, further research needs to be conducted that would allow more complex analyses between depression and offending. First, future research on depression and offending should present findings before and after controlling for confounds such as socioeconomic status, substance use, and previous offending. This will aid in determining whether depression remains predictive while accounting for potential shared risk factors. If depression no longer remains predictive, then this would suggest that the shared risk factors model provides the best explanatory model to understand the relationship.

Second, further research should investigate possible symptom-level relationships between depression and offending, such as examining whether an agitated depression (i.e., prominent irritability instead of low mood) would be more strongly related to risk for violence, compared to a melancholic depression (i.e., prominent symptoms of anhedonia), which may be unrelated to risk for offending (Gabbay et al., 2015). If irritable mood in the context of depression is predictive of offending, but not depressed mood, this will provide further support to the idea that irritability is the driving symptom in the acting out conceptualization of the depression – offending link.

Third, studies on mental illness and offending rarely assess the onset and course of an illness in combination with the course of offending. A design that tracks the onset and course of depression in combination with the trajectory of offending is exactly what is needed to make any conclusions regarding the specific reason that depression and offending co-occur. It is quite possible that the positive correlation between depression and offending may be due to other types of relationships besides a simple risk factor. In the current meta-analysis, only prospective studies were included thus ensuring that depression is correlated with *future* offending. However, it is unclear regarding whether offending could have occurred before depression and may have contributed to the onset of depression. For example, engaging in criminal behaviour and getting caught in particular may be a very stressful event that leads to rejection from family or peers. In turn, rejection may lead in part to the onset of depression or exacerbation of pre-existing depressive symptoms. Due to the rarity of studies controlling for previous offending, the current meta-analysis was not able to examine the existing research in such fine detail. Future research should be clear about indicating specific details about the timeline of depression and offending behaviour, in order to determine how these variables relate to each other.

Conclusion

In sum, the current review indicates that depression is associated with an increased risk for offending, particularly in community samples of adolescents. However, at the present moment, research is unable to determine *why* depression and offending are associated. Future research should ask more nuanced questions like what symptoms, under which circumstances and in combination with which personal factors is depression associated with offending. This goes beyond the simple correlational

research that typifies this area, to actually conduct theory-driven research that can further develop our understanding of how mental illness, such as depression, interacts with delinquency. This research is imperative to provide policymakers with evidence-based reasoning regarding how best to allocate funding and resources to ensure every adolescent has opportunities to live up to their full potential.

References

Cited Articles

Achenbach, T. M. (1990). *The Young Adult Self Report.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991a). *Manual for the Teacher's Report Form and 1991 profile*. Burlington VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology, 30*, 47-87. doi:10.1111/j.1745-9125.1992.tb01093.x

Agnew, R. (2006). General Strain Theory: Current status and directions for further research. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory.* (Vol. 15, pp. 101–123). Piscataway, NJ: Transaction Publishers. Retrieved from https://search-ebscohost-com.proxy.lib.sfu.ca/login.aspx?direct=true&db=psyh&AN=2006-22890-003&site=ehost-live. https://doi.org/10.4324/9781315130620-4

Ahrens, L., Rexford, J. (2002). Cognitive processing therapy for incarcerated adolescents with PTSD. In R. Greenwald (Ed.), *Trauma and juvenile delinquency: Theory, research and interventions.* Binghamton, NY: The Haworth Press. Retrieved from https://doi.org/10.1300/J146v06n01_10

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed). Washington, DC: Author. Retrieved from https://doi.org/10.1176/appi.books.9780890425596

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Cincinnati, OH: Anderson.

Angold, A., & Costello, E. J. (2000). The Child and Adolescent Psychiatric Assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(1), 39-48.

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*(4), 237–249.

Aske, J., Hale, B., Engels, R., Raaijmakers, Q., & Meeus, W. (2007). Co-occurrence of depression and delinquency in personality types. *European Journal of Personality: Published for the European Association of Personality Psychology*, *21*(2), 235-256. doi:10.1002/per.604

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation. Retrieved from https://doi.org/10.1037/t00742-000

Baerveldt, C., van Rossem, R., & Vermande, M. (2003). Pupils' delinquency and their social networks: A test of some network assumptions of the ability and inability models of delinquency. *The Netherlands Journal of Social Sciences, 39*, 107–125.

Birnbaum, A., Lytle, L. A., Murray, D. M., Story, M., Perry, C. L. & Boutelle, K. N. (2002). Survey development to assess correlates of young adolescents' eating. *American Journal of Health Behavior, 26*(4), 284-295. doi:10.5993/AJHB.26.4.5

Blitstein, J. L., Murray, D. M., Lytle, L. A., Birnbaum, A. S., & Perry, C. L. (2005). Predictors of violent behavior in an early adolescent cohort: Similarities and differences across genders. *Health Education and Behavior, 32*(2), 175-194. doi:10.1177/1090198104269516

Bonta, J., & Andrews, D. A. (2016). *The psychology of criminal conduct*. Retrieved from https://doi.org/10.4324/9781315677187

Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior*, *19*(3), 278–287. doi:10.1016/j.avb.2014.04.014

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*(2), 123-142. doi:10.1037/0033-2909.123.2.123

Boots, D. P., & Wareham, J. (2010). Does controlling for comorbidity matter? DSM-oriented scales and violent offending in Chicago youth. *Aggressive Behavior, 36*(3), 141-157. doi:10.1002/ab.20338

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis version 3*. Engelwood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* United Kingdom. Wiley. doi:10.1002/9780470743386

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R…. & advisory group (2009). *Comprehensive Meta Analysis user manual 3.0.* United Kingdom. Wiley.

Callaghan, J., Pace, F., Young, B., & Vostanis, P. (2003). Primary mental health workers within youth offending teams: A new service model. *Journal of Adolescence*, *26*(2), 185-199. doi:10.1016/S0140-1971(02)00131-8

Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up at Grade 8. *Development and Psychopathology, 4*, 125–144. doi:10.1017/S0954579400005605

Carswell, K., Maughan, B., Davis, H., Davenport, F., & Goddard, N. (2004). The psychosocial needs of young offenders and adolescents from an inner city area. *Journal of Adolescence, 27*(4), 415-428. doi:10.1016/j.adolescence.2004.04.003

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics Simulation and Computation, 39,* 860– 864. doi:10.1080/03610911003650383

Chitsabesan, P., Kroll, L., Bailey, S. U. E., Kenning, C., Sneider, S., MacDonald, W., & Theodosiou, L. (2006). Mental health needs of young offenders in custody and in the community. *The British Journal of Psychiatry*, *188*(6), 534-540.

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSMIV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy, 38*, 835–855. Doi:10.1016/S0005-7967(99)00130-8

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284. doi:10.1037/1040-3590.6.4.284

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, *70*(1), 6. doi:10.1037/0022-006X.70.1.6

Connell, A. M., & Dishion, T. J. (2006). The contribution of peers to monthly variation in adolescent depressed mood: A short-term longitudinal study with time-varying predictors. *Development and Psychopathology, 18*(1), 139-154. doi:10.1017/S0954579406060081

Costello, A., Edelbrock, C., Kalas, R., Kessler, M., Klaric, S.A. (1982). *The National Institute of Mental Health Diagnostic Interview Schedule for Children (DISC)*. Rockville, MD: National Institute of Mental Health.

Department of Justice. (2017, August 8). The Youth Criminal Justice Act Summary and Background. Retrieved from https://www.justice.gc.ca/eng/cj-jp/yj-jj/tools-outils/back-hist.html.

Derogatis, L.R. (1983) *SCL-90-R: administration, scoring and procedures manual-II.* Towson, MD: Clinical Psychometric Research.

Duran-Bonavila, S., Vigil-Colet, A., Cosi, S., & Morales-Vives, F. (2017). How individual and contextual factors affects antisocial and delinquent behaviors: A comparison between young offenders, adolescents at risk of social exclusion, and a community sample. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01825

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. doi: 10.1111/j.0006-341X.2000.00455.x

Egger, M., & Smith, G. D. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ: British Medical Journal (International Edition)*, *315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

Elliott, D.S.S., Huizinga, D., & Ageton, S.S. (1985). Explaining delinquency and drug use. Beverly Hills, CA: Sage Publications.

Elliott, D., Huizinga, D., & Menard, S. (1989). Multiple problem youth: Delinquency, substance use, and mental health problems. New York, NY: Springer-Verlag. doi:10.1007/978-1-4613-9637-6

Fagan, A. A., & Western, J. (2005). Escalation and deceleration of offending behaviours from adolescence to early adulthood. *Australian and New Zealand Journal of Criminology 38*(1), 59-76. doi:10.1375/acri.38.1.59

Farrington, D. P., Gallagher, B., Morley, L., & St. Ledger, R. J. (1988). Are there any successful men from criminogenic backgrounds? *Psychiatry: Interpersonal and Biological Processes*, *51*(2), 116-130. doi:10.1080/00332747.1988.11024387

Felson, R. B., Silver, E., & Remster, B. (2012). Mental disorder and offending in prison. *Criminal Justice and Behavior*, *39*(2), 125-143. doi:10.1177/0093854811428565

Gabbay, V., Johnson, A. R., Alonso, C. M., Evans, L. K., Babb, J. S., & Klein, R. G. (2015). Anhedonia, but not irritability, is associated with illness severity outcomes in adolescent major depression. *Journal of Child and Adolescent Psychopharmacology*, *25*(3), 194–200. doi:10.1089/cap.2014.0105

Garber, J., Clarke, G.N., Weersing, V.R., Beardslee, W.R., Brent, D.A., Gladstone, T.R., DeBar,… & Iyengar, S. (2009). Prevention of depression in at-risk adolescents: a randomized controlled trial. *Journal of the American Medical Association 301*, 2215–2224. doi:10.1001/jama.2009.788

Haddaway, N. R., Collins, A. M., Coughlin, D. & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLOS One, 10*(9). doi:10.1371/journal.pone.0138237

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3*, 339–353. doi:10.1037/1082-989X.3.3.339

Hein, S., Barbot, B., Square, A., Chapman, J., Geib, C. F., & Grigorenko, E. L. (2017). Violent offending among juveniles: A 7-year longitudinal study of recidivism, desistance, and associations with mental health. *Law and Human behavior*, *41*(3), 273-291. doi:10.1037/lhb0000241

Herrera, V. M. (2002). *Family influences on adolescent depression and delinquency: Gender differences in risk*. (Doctoral dissertation). Retrieved from ProQuest Information & Learning. (https://search-ebscohost-com.proxy.lib.sfu.ca/login.aspx?direct=true&db=psyh&AN=2002-95006-357&site=ehost-live)

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage. London, UK. doi:10.4135/9781483398105

Hodges, K., Kline, J., Fitch, P., McKnew, D. & Cytryn, L. (1981). The Child Assessment Schedule: a diagnostic interview for research and clinical use. *Catalog of Selected Documents in Psychology, 11,* 56-61. doi:10.1037/t06953-000

Holzer, K. J., Oh, S., Salas-Wright, C. P., Vaughn, M. G., & Landess, J. (2018). Gender differences in the trends and correlates of major depressive episodes among juvenile offenders in the United States. *Comprehensive Psychiatry*, *80*, 72–80. doi:10.1016/j.comppsych.2017.09.005

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11,* 193–206. doi:10.1037/1082-989X.11.2.193

Institute of Behavioral Science (1991) *Denver Youth Survey Youth Interview Schedule*. Boulder, CO: University of Colorado, Boulder.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124. doi:10.1371/journal.pmed.0020124

Jolliffe, D., Farrington, D. P., Piquero, A. R., MacLeod, J. F., & van de Weijer, S. (2017). Prevalence of life-course-persistent, adolescence-limited, and late-onset offenders: A systematic review of prospective longitudinal studies. *Aggression and Violent Behavior*, *33*, 4–14. doi:10.1016/j.avb.2017.01.009

Kandel, D. B., & Davies, M. (1982). Epidemiology of depressive mood in adolescents. *Archives of General Psychiatry, 39*, 1205–1212. doi:10.1001/archpsyc.1982.04290100065011

Katsiyannis, A., Zhang, D., Barrett, D. E., & Flaska, T. (2004). Background and psychosocial variables associated with recidivism among adolescent males: A 3-year investigation. *Journal of Emotional and Behavioral Disorders*, *12*(1), 23-29. doi:10.1177/10634266040120010301

Kenny, D. T., Lennings, C. J., & Nelson, P. K. (2007). The mental health of young offenders serving orders in the community: Implications for rehabilitation. *Journal of Offender Rehabilitation, 45*, 123–148. doi:10.1300/J076v45n01_10

Kerig, P. K., & Becker, S. P. (2012). Trauma and girls' delinquency. In *Delinquent girls* (pp. 119-143). New York, NY: Springer. doi:10.1007/978-1-4614-0415-6_8

Kerig, P., & Schindler, S. (2013). Engendering the evidence base: A critical review of the conceptual and empirical foundations of gender-responsive interventions for girls' delinquency. *Laws*, *2*(3), 244-282.

Kerr, M., Tremblay, R. E., Pagani, L., & Vitaro, F. (1997). Boys' behavioral inhibition and the risk of later delinquency. *Archives of General Psychiatry*, *54*(9), 809-816. doi:10.1001/archpsyc.1997.01830210049005

Kessler, R.C., Berglund, P. Demler, O., et al (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA 289*(23), 3095–3105. doi:10.1001/jama.289.23.3095

Kicinski, M., Springate, D. A., & Kontopantelis, E. (2015). Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. *Statistics in Medicine*, *34*(20), 2781-2793. https://doi.org/10.1002/sim.6525

Kiesner, J. (2002). Depressive' symptoms in early adolescence: Their relations with classroom problem' behavior and peer status. *Journal of Research on Adolescence, 12*, 463-478. doi:10.1111/1532-7795.00042

Kim, H., & Kim, H. (2005). Gender differences in delinquent behavior among Korean adolescents. *Child Psychiatry and Human Development, 35*(4), 325-345. doi:10.1007/s10578-005-2691-1

Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin*, *21*, 995-998.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (1999). Patient Health Questionnaire-9. *PsycTESTS*. doi:10.1037/t06165-000

Langhinrichsen-Rohling, J., & Lamis, D. A. (2008). Current suicide proneness and past suicidal behavior in adjudicated adolescents. *Suicide and Life-Threatening Behavior*, *38*(4), 415–426. doi:10.1521/suli.2008.38.4.415

Loeber, R., & Keenan, K. (1994). Interaction between conduct disorder and its comorbid conditions: Effects of age and gender. *Clinical Psychology Review*, *14*(6), 497–523. doi:10.1016/0272-7358(94)90015-9

Luchini, C., Stubbs, B., Solmi, M., & Veronese, N. (2017). Assessing the quality of studies in meta-analyses: Advantages and limitations of the Newcastle Ottawa Scale. *World Journal of Meta-Analysis. 5*(4): 80-84. doi:10.13105/wjma.v5.i4.80

McCormick, S., Peterson-Badali, M., & Skilling, T. A. (2015). Mental health and justice system involvement: A conceptual analysis of the literature. *Psychology, Public Policy, and Law, 21*(2), 213–225. doi:10.1037/law0000033

McCormick, S., Peterson-Badali, M., & Skilling, T. A. (2017). The role of mental health and specific responsivity in juvenile justice rehabilitation. *Law and Human Behavior*, *41*(1), 55–67. doi:10.1037/lhb0000228

Messer, S. C., Angold, A., Costello, E. J., Loeber, R., Van Kammen, W., & Stouthamer-Loeber, M. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: Factor composition and structure across development. *International Journal of Methods in Psychiatric Research*, *5*, 251-262.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Morey, L. C. (2007). *Personality Assessment Inventory-Adolescent (PAI-A)*. Lutz, FL: Psychological Assessment Resources.

Morgan, R. D., Gendreau, P., Smith, P., Gray, A. L., Labrecque, R. M., MacLean, N., … Mills, J. F. (2016). Quantitative synthesis of the effects of administrative segregation on inmates' well-being. P*sychology, Public Policy, and Law, 22*, 439-461. doi:10.1037/law0000089

Obeidallah, D., Brennan, R. T., Brooks-Gunn, J., & Earls, F. (2004). Links between pubertal timing and neighborhood contexts: Implications for girls' violent behavior. *Journal of The American Academy of Child and Adolescent Psychiatry, 43*(12), 1460-1468. doi:10.1097/01.chi.0000142667.52062.1e

Perkins, K., Ferrari, N., Rosas, A., Bessette, R., Williams, A., & Omar, H. (1997). You won't know unless you ask: The biopsychosocial interview for adolescents. *Clinical Pediatrics*, *36*(2), 79-86. doi:10.1177/000992289703600204

Pfeffer, C. R., Plutchik, R., & Mizruchi, M. S. (1983). Predictors of assaultiveness in latency age children. *The American Journal of Psychiatry, 140*(1), 31-35. doi:10.1176/ajp.140.1.31

Piquero, N. L., & Sealock, M. D. (2000). Generalizing general strain theory: An examination of an offending population. *Justice Quarterly*, *17*(3), 449-484. doi:10.1080/07418820000094631

Poznanski, E. O., Cook, S. C., & Carroll, B. J. (1979). A depression rating scale for children. *Pediatrics,1001*, 48109.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385-401. doi:10.1177/014662167700100306

Reiter, A. M., Suzuki, S., O'Doherty, J. P., Li, S. C., & Eppinger, B. (2019). Risk contagion by peers affects learning and decision-making in adolescents. *Journal of Experimental Psychology: General*. *148*(9), 1494–1504. doi:10.1037/xge0000512

Reynolds, W. M. (1987). *Reynolds Adolescent Depression Scale (RADS– 2).* Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M. (2002). *Reynolds Adolescent Depression Scale: Professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Ritakallio, M., Kaltiala-Heino, R., Kivivuori, J., & Rimpelä, M. (2005). Brief report: Delinquent behaviour and depression in middle adolescence: A Finnish community sample. *Journal of Adolescence, 28*(1), 155-159. doi :10.1016/j.adolescence.2004.07.002

Ritakallio, M., Koivisto, A., von der Pahlen, B., Pelkonen, M., Marttunen, M., & Kaltiala-Heino, R. (2008). Continuity, comorbidity and longitudinal associations between depression and antisocial behaviour in middle adolescence: A 2-year prospective follow-up study. *Journal of Adolescence, 31*(3), 355-370. doi :10.1016/j.adolescence.2007.06.006

Robins, L.N., Helzer, J.E., Croughan, J., Williams, J.B.W., Spitzer, R.L. (1981). *NIMH Diagnostic Interview Schedule: Version III*. Rockville, MD: National Institute of Mental Health

Roseman, M., Kloda, L.A., & Saadat, N. (2016). Accuracy of depression screening tools to detect major depression in children and adolescents: a systematic review. *Canadian Journal of Psychiatry*, *61*(12):746–757. doi:10.1177/0706743716651833

Roseman, M., Saadat, N., Riehm, K. E., Kloda, L. A., Boruff, J., Ickowicz, A., … Thombs, B. D. (2017). Depression screening and health outcomes in children and adolescents: A systematic review. *The Canadian Journal of Psychiatry / La Revue Canadienne de Psychiatrie*, *62*(12), 813–817. doi:10.1177/0706743717727243

Ruchkin, V. V., Schwab-Stone, M., Koposov, R. A., Vermeiren, R., & King, R. A. (2003). Suicidal ideations and attempts in juvenile delinquents. *Journal of Child Psychology and Psychiatry*, *44*(7), 1058–1066. doi:10.1111/1469-7610.0019

Rutter, M., & Sroufe, L. A. (2000). Developmental psychopathology: Concepts and challenges. *Development and Psychopathology*, *12*(3), 265-296. doi:10.1017/S0954579400003023

Schuch, J. J., Roest, A. M., Nolen, W. A., Penninx, B. H., & de Jonge, P. (2014). Gender differences in major depressive disorder: Results from the Netherlands study of depression and anxiety. *Journal of Affective Disorders*, *156*, 156-163. doi:10.1016/j.jad.2013.12.011

Scott, T., & Brown, S. L. (2018). Risks, strengths, gender, and recidivism among justice-involved youth: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *86*(11), 931–945. doi:10.1037/ccp0000343

Shaffer, D. (1998). Epidemiological aspects of some problems in child and adolescent psychiatry. *Epidemiologia e Psichiatria Sociale*, *7*(3), 151–155. doi:10.1017/S1121189X00007326

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(1), 28-38. doi:10.1097/00004583-200001000-00014

Sigfusdottir, I., Farkas, G., & Silver, E. (2004). The role of depressed mood and anger in the relationship between family conflict and delinquent behavior. *Journal of Youth and Adolescence, 33*(6), 509-522. doi:10.1023/B:JOYO.0000048065.17118.63

Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology, 25*(9), 603-605. doi:10.1007/s10654-010-9491-z

Steer, R.A., Cavalieri, T.A., Leonard, D.M., Beck, A.T. (1999) Use of the Beck Depression Inventory for Primary Care to screen for major depression disorders. *General Hospital Psychiatry*, *21*, 106–11. doi:10.1016/S0163-8343(98)00070-X

Straus, M. A., & Douglas, E. M. (2004). A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence and Victims*, *19*(5), 507-520. doi:10.1891/vivi.19.5.507.63686

Straus, M. A. (1979). Family patterns and child abuse in a nationally representative American sample. *Child Abuse and Neglect*, *3*(1), 213-225. doi:10.1016/0145-2134(79)90034-6

Stringaris, A., Maughan, B., Copeland, W. S., Costello, E. J., & Angold, A. (2013). Irritable mood as a symptom of depression in youth: prevalence, developmental, and clinical correlates in the Great Smoky Mountains Study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(8), 831-840. doi:10.1016/j.jaac.2013.05.017

Taylor, D. C., & Ounsted, C. (1972). The nature of gender differences explored through ontogenetic analysis of sex ratios in disease. In C. Ounsted & D. C. Taylor (Eds.). *Gender differences: Their ontogeny and significance* (pp. 21 - 240). London: Churchill Livingstone.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112-125. doi: 10.1002/jrsm.11

Viljoen, J., Nicholls, T., Cruise, K., Benetau-Douglas, J., Desmarais, S., & Barone, C., & Petersen, K., … & Webster, C. (2016). START:AV Knowledge Guide- A research compendium on the START:AV strength and vulnerability items. Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute.

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., […], & Losos, M.(N.D). The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/ oxford.htm

White, J. (2014). *Practice guidelines for working with children and youth at-risk for suicide in community mental health settings*. Ministry of Children and Family Development. Retrieved from https://www2.gov.bc.ca/assets/gov/health/managing-your-health/mental-health-substance-use/child-teen-mental-health/practice_guidelines_children_youth_at_risk_suicide.pdf

Wiens, K., Williams, J. V., Lavorato, D. H., Duffy, A., Pringsheim, T. M., Sajobi, T. T., & Patten, S. B. (2017). Is the prevalence of major depression increasing in the Canadian adolescent population? Assessing trends from 2000 to 2014. *Journal of Affective Disorders*, *210*, 22-26. doi:10.1016/j.jad.2016.11.018

Wiesner, M., & Kim, H. K. (2006). Co-occurring delinquency and depressive symptoms of adolescent boys and girls: A dual trajectory modeling approach. *Developmental Psychology, 42*(6), 1220-1235. doi:10.1037/0012-1649.42.6.1220

Wolff, J. C., & Ollendick, T. H. (2006). The comorbidity of conduct problems and depression in childhood and adolescence. *Clinical Child and Family Psychology Review*, *9*(3–4), 201–220. doi:10.1007/s10567-006-0011-3.

World Health Organization. (1986). *Young people's health-a challenge for society: report of a WHO study group on young people and health for all by the year 2000. Retrieved from* https://www.who.int/maternal_child_adolescent/documents/pdfs/9241591269_op_handout.pdf

Yu, R., Aaltonen, M., Branje, S., Ristikari, T., Meeus, W., Salmela-Aro, K., … Fazel, S. (2017). Depression and violence in adolescence and young adults: Findings from three longitudinal cohorts. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(8), 652–658. doi:10.1016/j.jaac.2017.05.016

Studies Included in the Meta-analysis

Anderson, D. M. (2015). Youth depression and future criminal behavior. *Economic Inquiry*, *53*(1), 294–317. doi:10.1111/ecin.12145

Aske, J., Hale, B., Engels, R., Raaijmakers, Q., & Meeus, W. (2007). Co-occurrence of depression and delinquency in personality types. *European Journal of Personality: Published for the European Association of Personality Psychology*, *21*(2), 235-256. doi:10.1002/per.604

Beyers, J. M., & Loeber, R. (2003). Untangling developmental relations between depressed mood and delinquency in male adolescents. *Journal of Abnormal Child Psychology*, *31*(3), 247–266. doi:10.1023/A:1023225428957

Blitstein, J. L., Murray, D. M., Lytle, L. A., Birnbaum, A. S., & Perry, C. L. (2005). Predictors of violent behavior in an early adolescent cohort: Similarities and differences across genders. *Health Education and Behavior*, *32*(2), 175-194. doi:10.1177/1090198104269516

Capaldi, D. M., & Stoolmiller, M. (1999). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: III. Prediction to young-adult adjustment. *Development and Psychopathology*, *11*(1), 59-84. doi:10.1017/S0954579499001959

Caprara, G. V., Gerbino, M., Paciello, M., Di Giunta, L., & Pastorelli, C. (2010). Counteracting depression and delinquency in late adolescence: The role of regulatory emotional and interpersonal self-efficacy beliefs. *European Psychologist*, *15*(1), 34–48. doi:10.1027/1016-9040/a000004

Cochrane, D., & Viljoen, J. (2016*). Untangling the relationship between anxiety, depression, and offending* (Unpublished honours thesis). Simon Fraser University, Burnaby, Canada.

Copeland, W. E., Miller-Johnson, S., Keeler, G., Angold, A., & Costello, E. J. (2007). Childhood psychiatric disorders and young adult crime: A prospective, population-based study. *The American Journal of Psychiatry*, *164*(11), 1668–1675. doi:10.1176/appi.ajp.2007.06122026

Diamantopoulou, S., Verhulst, F. C., & van der Ende, J. (2011). Gender differences in the development and adult outcome of co-occurring depression and delinquency in adolescence. *Journal of Abnormal Psychology*, *120*(3), 644–655. doi:10.1037/a0023669

Elkington, K. S., Teplin, L. A., Abram, K. M., Jakubowski, J. A., Dulcan, M. K., & Welty, L. J. (2015). Psychiatric disorders and violence: a study of delinquent youth after detention. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*(4), 302-312. doi: 10.1016/j.jaac.2015.01.002

Herrera, V. M. (2002). *Family influences on adolescent depression and delinquency: Gender differences in risk*. (Doctoral dissertation). Retrieved from ProQuest Information & Learning. (https://search-ebscohost-com.proxy.lib.sfu.ca/login.aspx?direct=true&db=psyh&AN=2002-95006-357&site=ehost-live)

Katsiyannis, A., Zhang, D., Barrett, D. E., & Flaska, T. (2004). Background and psychosocial variables associated with recidivism among adolescent males: A 3-year investigation. *Journal of Emotional and Behavioral Disorders*, *12*(1), 23-29. doi:10.1177/10634266040120010301

Mason, W. A., Hitchings, J. E., & Spoth, R. L. (2007). Emergence of delinquency and depressed mood throughout adolescence as predictors of late adolescent problem substance use. *Psychology of Addictive Behaviors*, *21*(1), 13. doi:10.1037/0893-164X.21.1.13

Mason, W. A., Kosterman, R., Hawkins, J. D., Herrenkohl, T. I., Lengua, L. J., & McCauley, E. (2004). Predicting depression, social phobia, and violence in early adulthood from childhood behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(3), 307–315. doi:10.1097/00004583-200403000-00012

McCarty, C. A., Mason, W. A., Kosterman, R., Hawkins, J. D., Lengua, L. J., & McCauley, E. (2008). Adolescent school failure predicts later depression among girls. *Journal of Adolescent Health*, *43*(2), 180-187. doi:10.1016/j.jadohealth.2008.01.023

Moon, B., Morash, M., McCluskey, C. P., & Hwang, H.-W. (2009). A comprehensive test of general strain theory: Key strains, situational- and trait-based negative emotions, conditioning factors, and delinquency. *Journal of Research in Crime and Delinquency*, *46*(2), 182–212. doi:10.1177/0022427808330873

Ostrowsky, M. K. (2007). *Extending Khantzian's self-medication hypothesis: An examination of low self-esteem, depression, alcohol use, marijuana use, and violent behavior*. (Doctoral dissertation). Retrieved from ProQuest Information & Learning. (https://search-ebscohost com.proxy.lib.sfu.ca/login.aspx?direct=true&db=psyh&AN=2007-99015-033&site=ehost-live)

Overbeek, G., Biesecker, G., Kerr, M., Stattin, H., Meeus, W., & Engels, R. C. (2006). Co-occurrence of depressive moods and delinquency in early adolescence: The role of failure expectations, manipulativeness, and social contexts. *International Journal of Behavioral Development*, *30*(5), 433-443. doi:10.1177/0165025406071491

Pardini, D. A., Loeber, R., Farrington, D. P., & Stouthamer-Loeber, M. (2012). Identifying direct protective factors for nonviolence. *American Journal of Preventive Medicine*, *43*(2), S28–S40. doi:10.1016/j.amepre.2012.04.024

Park, A. (2012). *The impact of exposure to domestic violence on developmental trajectories of depressive symptoms and antisocial behavior across the transition to adulthood*. (Doctoral dissertation). Retrieved from: ProQuest Information & Learning. (https://search-ebscohost-com.proxy.lib.sfu.ca/login.aspx?direct=true&db=psyh&AN=2012-99090-523&site=ehost-live)

Siennick, S. E. (2007). The timing and mechanisms of the offending--depression link. *Criminology: An Interdisciplinary Journal*, *45*(3), 583–615. doi:10.1111/j.1745-9125.2007.00091.x

Simons, R. L., Chen, Y., Stewart, E. A., & Brody G. H. (2003). Incidents of discrimination and risk for delinquency: A longitudinal test of strain theory with an African American sample. *Justice Quarterly*, *20*(4), 827-854. doi:10.1080/07418820300095711

Stuewig, J., & McCloskey, L. A. (2005). The relation of child maltreatment to shame and guilt among adolescents: psychological routes to depression and delinquency. *Child Maltreatment*, *10*(4), 324–336. doi:10.1177/1077559505279308

Thomas, A. G., Ozbardakci, N., Fine, A., Steinberg, L., Frick, P. J., & Cauffman, E. (2018). Effects of physical and emotional maternal hostility on adolescents' depression and reoffending. *Journal of Research on Adolescence*, *28*(2), 427–437. doi:10.1111/jora.12345

Vieno, A., Kiesner, J., Pastore, M., & Santinello, M. (2008). Antisocial behavior and depressive symptoms: Longitudinal and concurrent relations. *Adolescence*, *43*(171), 649.

Wareham, J., & Boots, D. (2012). The link between mental health problems and youth violence in adolescence: A multilevel test of DSM-Oriented Problems. *Criminal Justice and Behavior, 39*(8), 1003-1024. doi:10.1177/0093854812439290

Weaver, C. M., Borkowski, J. G., & Whitman, T. L. (2008). Violence breeds violence: Childhood exposure and adolescent conduct problems. *Journal of Community Psychology*, *36*(1), 96-112. doi:10.1002/jcop.20219

You, S., & Lim, S. A. (2015). Development pathways from abusive parenting to delinquency: the mediating role of depression and aggression. *Child Abuse and Neglect*, *46*, 152–162. doi:10.1016/j.chiabu.2015.05.009

Yu, R., Aaltonen, M., Branje, S., Ristikari, T., Meeus, W., Salmela-Aro, K., … Fazel, S. (2017). Depression and violence in adolescence and young adults: Findings from three longitudinal cohorts. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(8), 652–658.e1. doi:10.1016/j.jaac.2017.05.016

Appendix A.

Depression and Offending Meta-Analysis Coding Manual

Inclusion Criteria

- Empirical study
- Published or disseminated in English
- Prospective design
- Includes a sample of offenders
- Includes a measure of depression
- Includes a measure of offending (general, violent)

Multiple Studies Reporting on a Single Study

In the event that there are multiple studies reporting on a single sample or overlapping samples, studies should be prioritized according to the following (taking sample size and comprehensiveness into consideration):

- Journal Article
- Dissertation
- Book Chapter
- Government Report
- Conference Presentation

Study Number : _____.

Study Title : _____.

Author(s) : _____.

Year : _____.

Rater: _____.

Study Characteristics

Type of Publication
1 = Journal Article
2 = Dissertation
3 = Book Chapter
4 = Government Report
5 = Conference Presentation
1 = other**:** _____.

Publication Bias (Peer Review)
0 = peer reviewed
1 = not peer reviewed

Type of Sample
0 = correctional
1 = forensic psychiatric
2 = probation
3 = non-offender
2 = other: _____.

Sample Gender (percent male):
0 = male only
1 = female only
2 = mixed
3 = unknown

Sample Age (mean age and SD):
0 = adolescent
1 = adult
2 = mixed

Country of Origin
0 = Canada
1 = United States
2 = Sweden
3 = Netherlands
4 = UK
5 = Australia
6 = New Zealand
7 = Other: _____.

Sample Size
Largest *n* with offending info **=** : _____.

Measure of Depression Information

Measure Name	Mean	SD	Range
1			
2			

Type of Measure of Depression

0 = official diagnosis with DSM or ICD by clinician
1 = self report psychological test
2 = behaviour coding
3 = other

Code for each measure (circle one)

Measure		
1 =	0	1
2 =	0	1

Describe the measure (i.e. what symptoms are covered):

1)

2)

Recidivism Information

Charges/Convictions
0 = convictions only
1 = charges only
2 = charges and convictions
3 = self report
4 = other: _____.

Average length of follow up (in months) = :

.

Fixed Follow Up
0 = no (follow up length varied between participants)
1 = yes (follow up was the same length for every participant)

If Fixed follow up how long?

Effect Size Information

Effect Size Table – Any Offending
Initial onset or recidivism (circle one)

Measure	*n* (base rate)	AUC (SE)	r_{pb} (*d*)	OR (SE)	Recidivists *M*(SD)	Non-Recidivists *M*(SD)
	n =	AUC=	$r_{pb=}$	OR=	*M*=	*M*=
	Base rate=	SE=	*d*=	SE=	SD=	SD=
	n =	AUC=	$r_{pb=}$	OR=	*M*=	*M*=
	Base rate=	SE=	*d*=	SE=	SD=	SD=

Did they control for any variables? If so describe the variables and what they found after the variables were controlled:

Effect Size Table – Violent Offending
Initial onset or recidivism (circle one)

Measure	*n* (base rate)	AUC (SE)	r_{pb} (*d*)	OR (SE)	Recidivists *M*(SD)	Non-Recidivists *M*(SD)
	n =	AUC=	$r_{pb=}$	OR=	*M*=	*M*=
	Base rate=	SE=	*d*=	SE=	SD=	SD=
	n =	AUC=	$r_{pb=}$	OR=	*M*=	*M*=
	Base rate=	SE=	*d*=	SE=	SD=	SD=

Did they control for any variables? If so describe the variables and what they found after the variables were controlled:

Effect Size Table – Other:
Initial onset or recidivism (circle one)

Measure	*n* (base rate)	AUC (SE)	r_{pb} (*d*)	OR (SE)	Recidivists *M*(SD)	Non-Recidivists *M*(SD)
	n =	AUC=	$r_{pb=}$	OR=	*M*=	*M*=
	Base rate=	SE=	*d*=	SE=	SD=	SD=
	n =	AUC=	$r_{pb=}$	OR=	*M*=	*M*=
	Base rate=	SE=	*d*=	SE=	SD=	SD=

Did they control for any variables? If so describe the variables and what they found after the variables were controlled:

Appendix B.

Modified Version of the Newcastle-Ottawa Quality Assessment Scale Cohort Studies

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection /3 possible stars

1) Representativeness of the exposed cohort
a) truly representative of the average _____ (describe) in the community ¯
b) somewhat representative of the average _____ in the community ¯
c) selected group of users eg nurses, volunteers
d) no description of the derivation of the cohort

2) Selection of the non exposed cohort
a) drawn from the same community as the exposed cohort ¯
b) drawn from a different source
c) no description of the derivation of the non exposed cohort

3) Ascertainment of exposure
a) secure record (eg surgical records) ¯
b) structured interview ¯
c) written self report
d) no description

Comparability /2 possible stars

1) Comparability of cohorts on the basis of the design or analysis ¯
a) study controls for _____ (select the most important factor) ¯
b) study controls for any additional factor ¯ (This criteria could be modified to indicate specific
control for a second important factor.)

Outcome /3 possible stars

1) Assessment of outcome
a) independent blind assessment ¯
b) record linkage ¯
c) self report
d) no description

2) Was follow-up long enough for outcomes to occur
a) yes (select an adequate follow up period for outcome of interest) ¯
b) no

3) Adequacy of follow up of cohorts
a) complete follow up - all subjects accounted for ¯
b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an
adequate %) follow up, or description provided of those lost) ¯
c) follow up rate < ____% (select an adequate %) and no description of those lost
d) no statement

Total Score /8 stars
Describe reasoning:

Appendix C.

Forest Plots

Study name	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value
02 - Aske et al. (2007)	1.290	0.874	1.904	1.281	0.200
03 - Beyers & Loeber (2003)	1.427	1.038	1.963	2.189	0.029
05 - Capaldi & Stoolmiller (1999)	2.018	1.206	3.376	2.674	0.008
06 - Caprara et al. (2010)	1.035	0.737	1.452	0.197	0.844
07 - Cochrane et al. (2016)	1.441	0.797	2.607	1.208	0.227
08 - Copeland et al. (2007)	2.300	0.778	6.803	1.505	0.132
09 - Diamantopoulou et al. (2011)	1.075	0.877	1.319	0.696	0.486
11 - Herrera (2001)	1.526	1.001	2.326	1.965	0.049
12 - Katsiyannia et al. (2004)	0.631	0.416	0.957	-2.168	0.030
14 - Mason et al. (2007)	2.178	1.526	3.107	4.292	0.000
15 - McCarty et al. (2008)	1.309	1.018	1.685	2.097	0.036
18 - Overbeek et al. (2006)	1.942	1.538	2.453	5.571	0.000
20 - Park (2012)	1.552	1.221	1.972	3.596	0.000
21 - Siennick (2007)	1.029	0.968	1.095	0.917	0.359
22 - Simons et al. (2003)	4.813	3.579	6.473	10.395	0.000
23 - Stuewig & McCloskey (2005)	1.314	0.855	2.018	1.246	0.213
24 - Thomas et al. (2017)	2.452	1.984	3.030	8.300	0.000
25 - Vieno et al. (2008)	5.628	2.600	12.182	4.385	0.000
27 - Weaver et al. (2008)	0.901	0.410	1.981	-0.259	0.795
	1.582	1.268	1.972	4.068	0.000

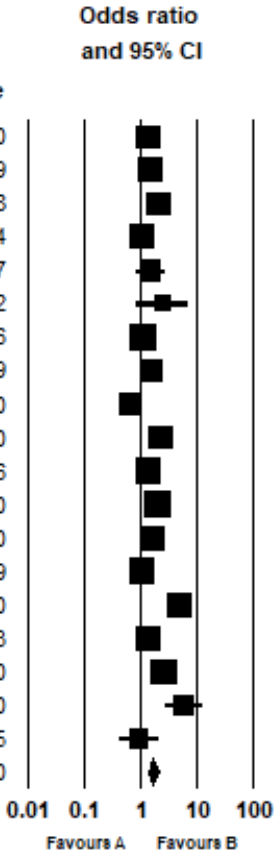

Figure C.1 Forest Plot: Odds of any offending

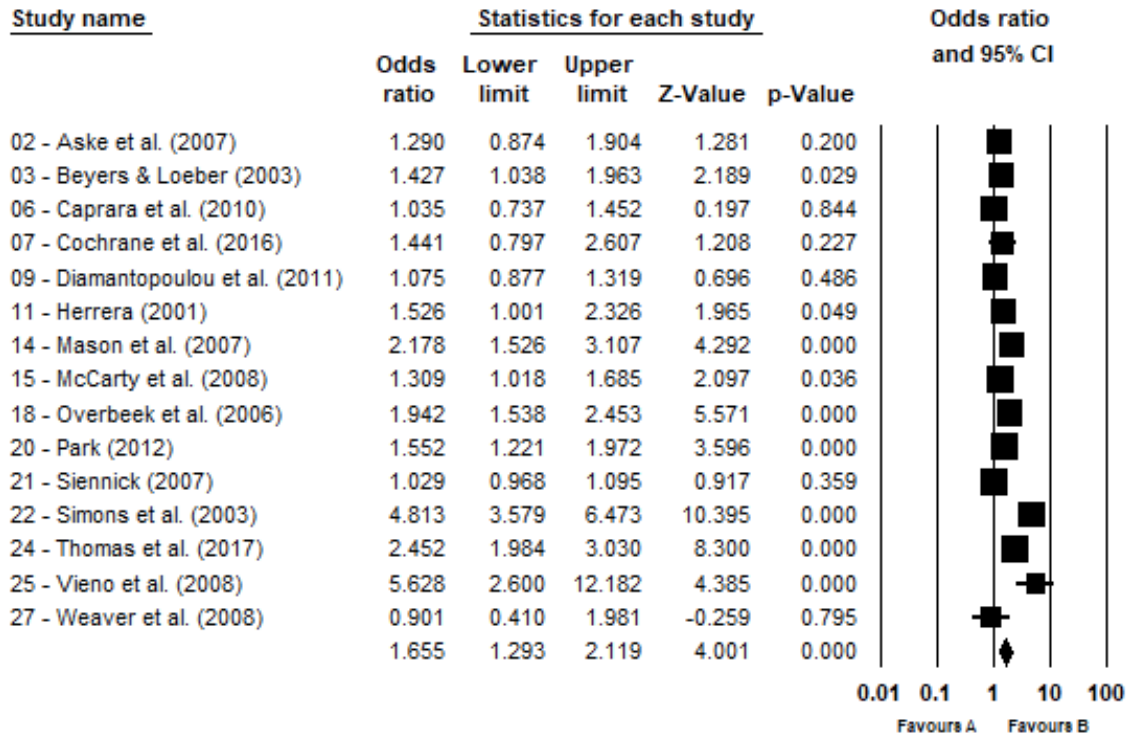

Figure C.2 Forest Plot: Odds of any offending measured by self-report

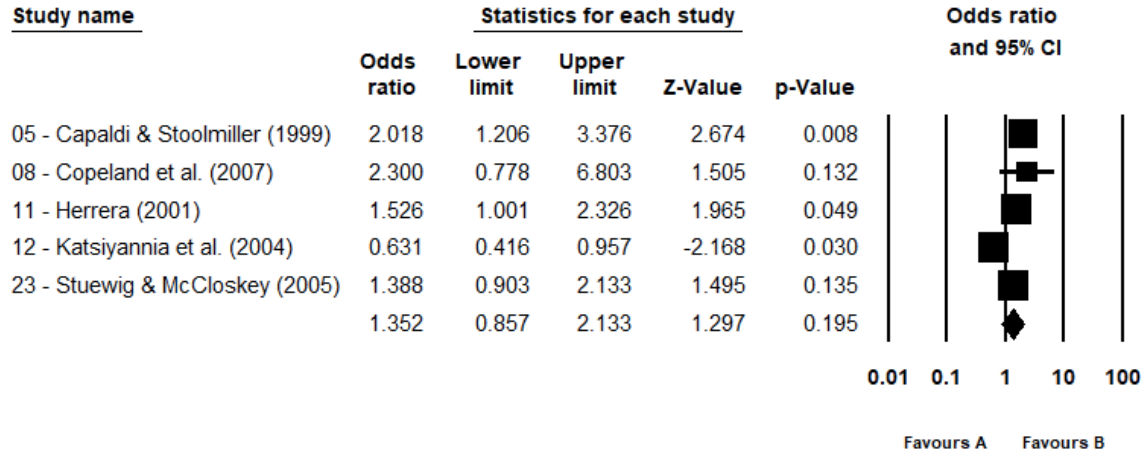

Figure C.3 Forest Plot: Odds of any offending as measured by official records

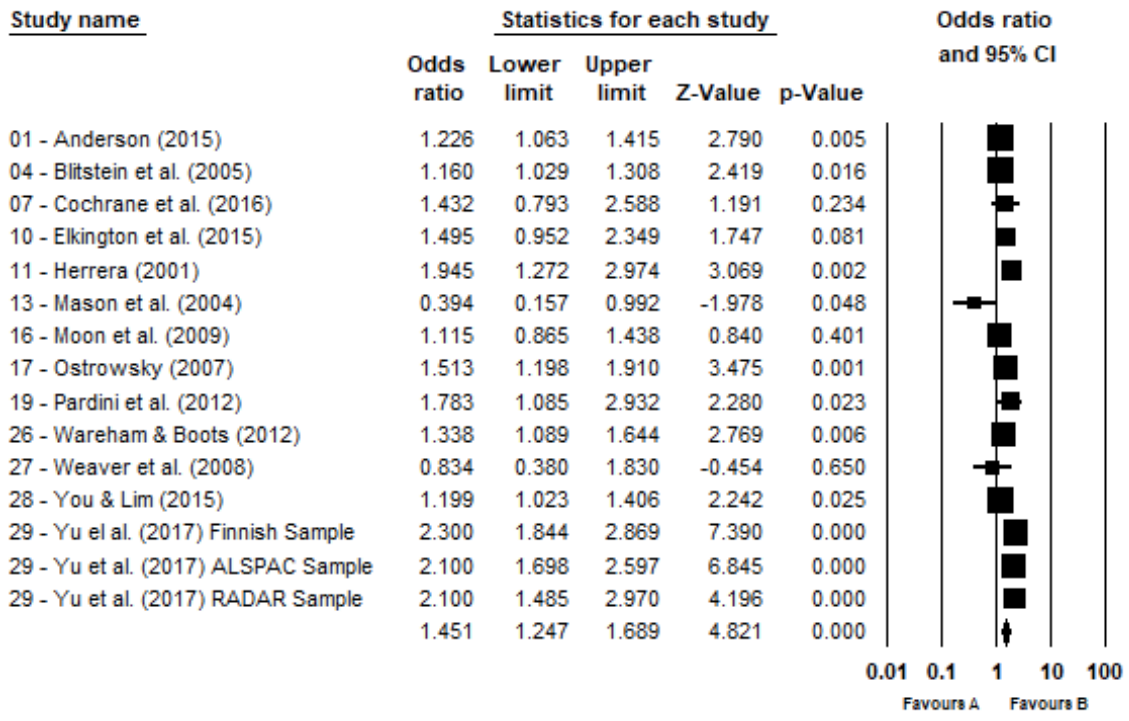

Figure C.4 Forest Plot: Odds of violent offending

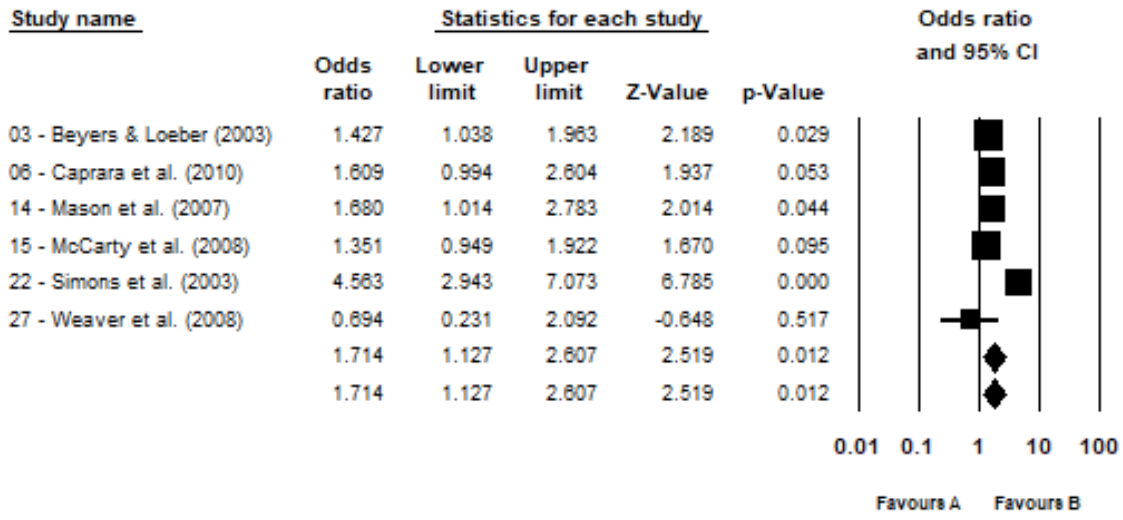

Figure C.5 Forest Plot: Odds of any offending for males

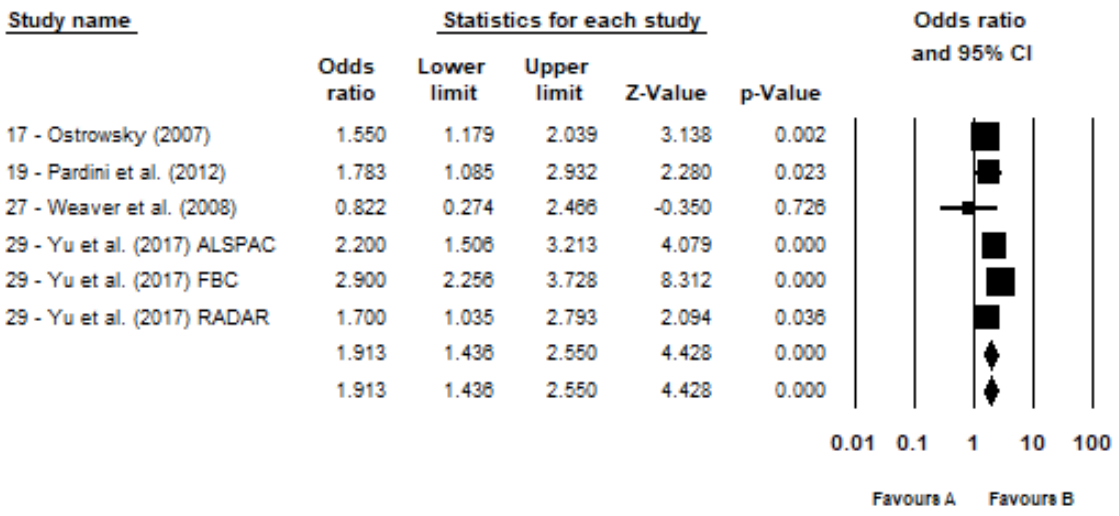

Figure C.6 Forest Plot: Odds of violent offending for males

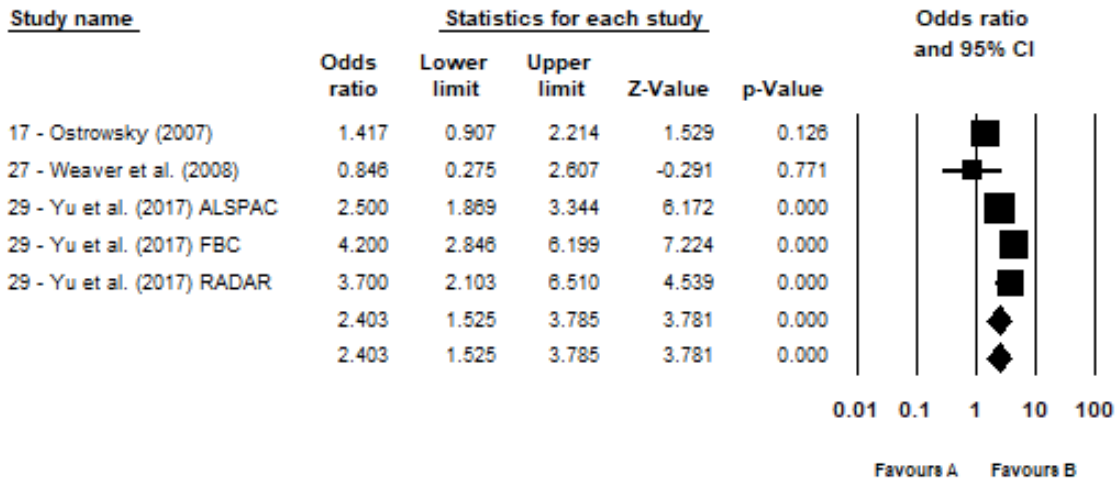

Figure C.7 Forest Plot: Odds of any offending for females

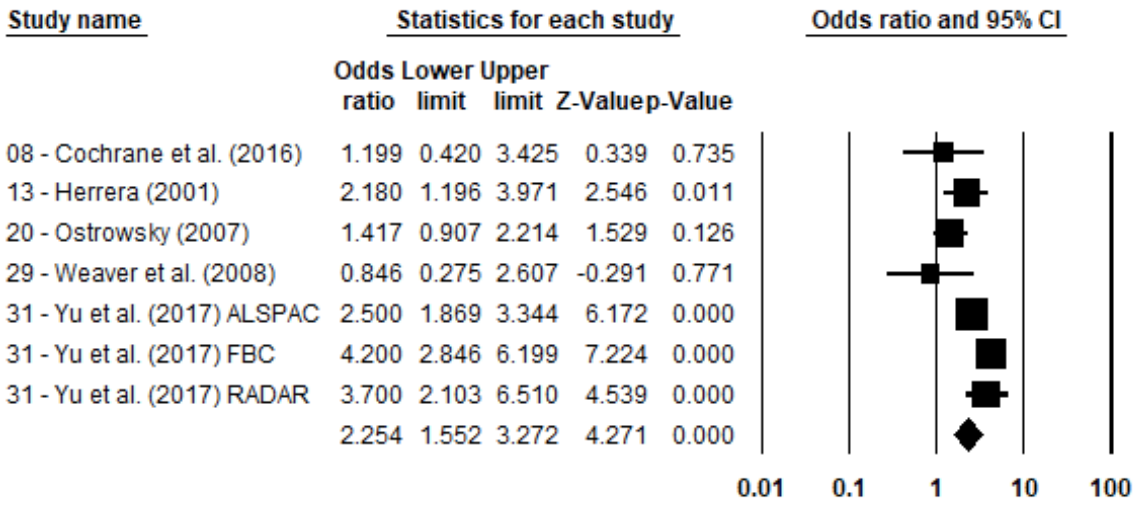

Figure C.8 Forest Plot: Odds of violent offending for females

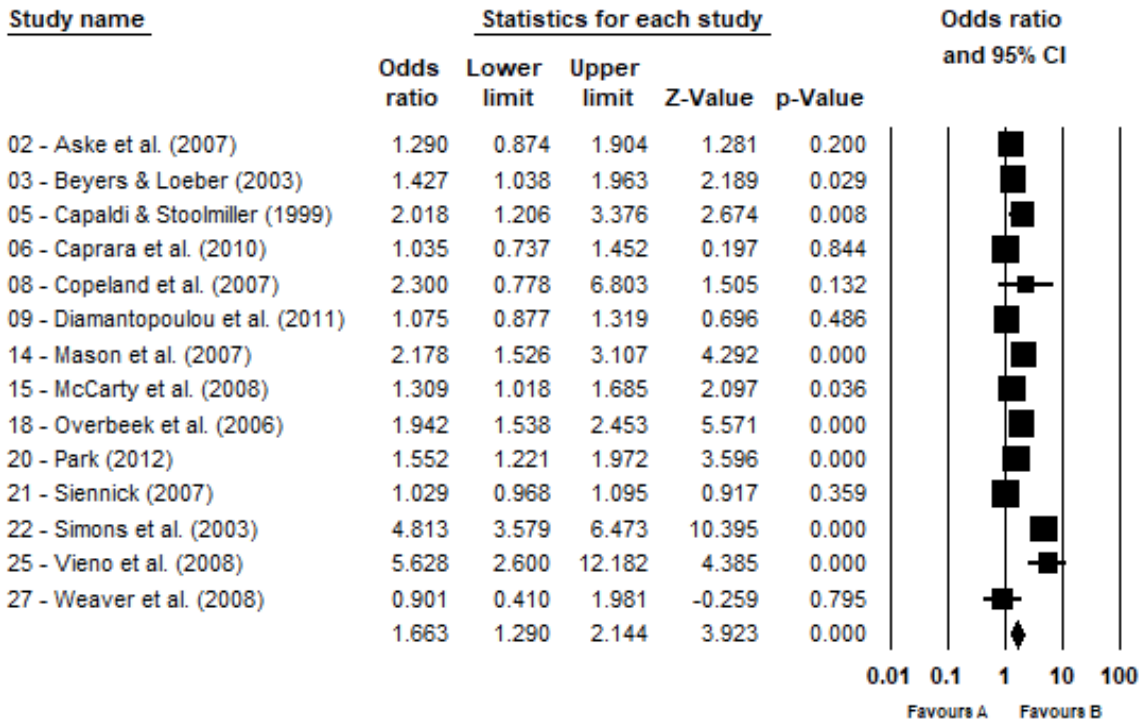

Figure C.9 Forest Plot: Odds of any offending in community samples

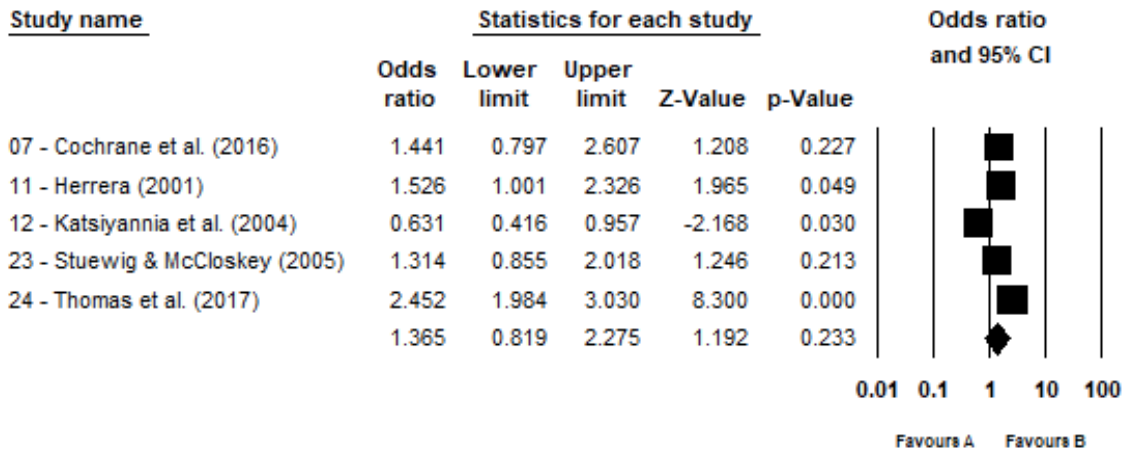

Figure C.10 Forest Plot: Odds of any offending in justice-involved samples

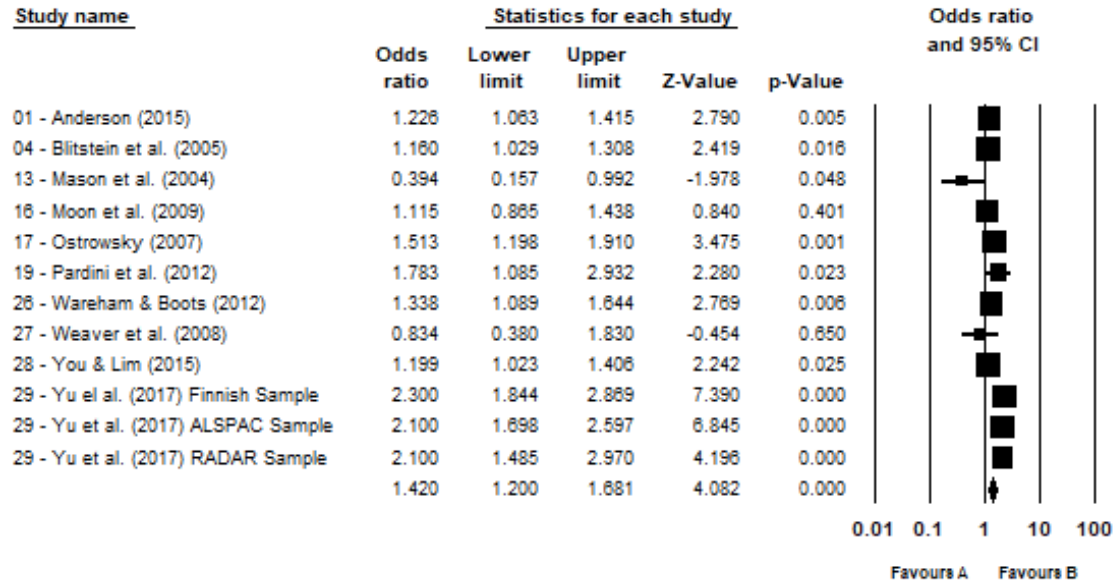

Figure C.11 Forest Plot: Odds of violent offending in community samples

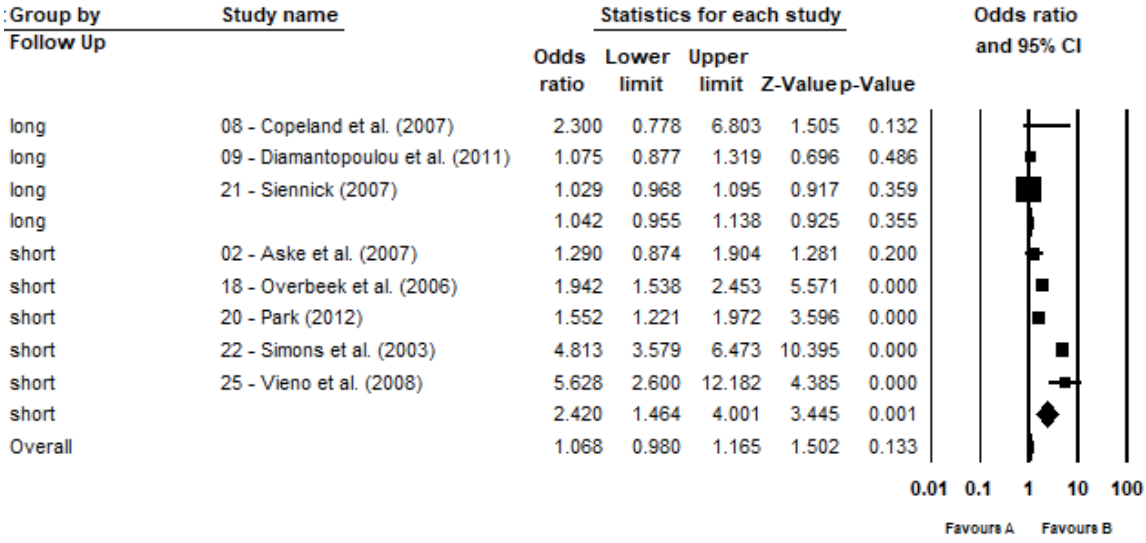

Figure C.12 Forest Plot: Odds of any offending by follow up time length

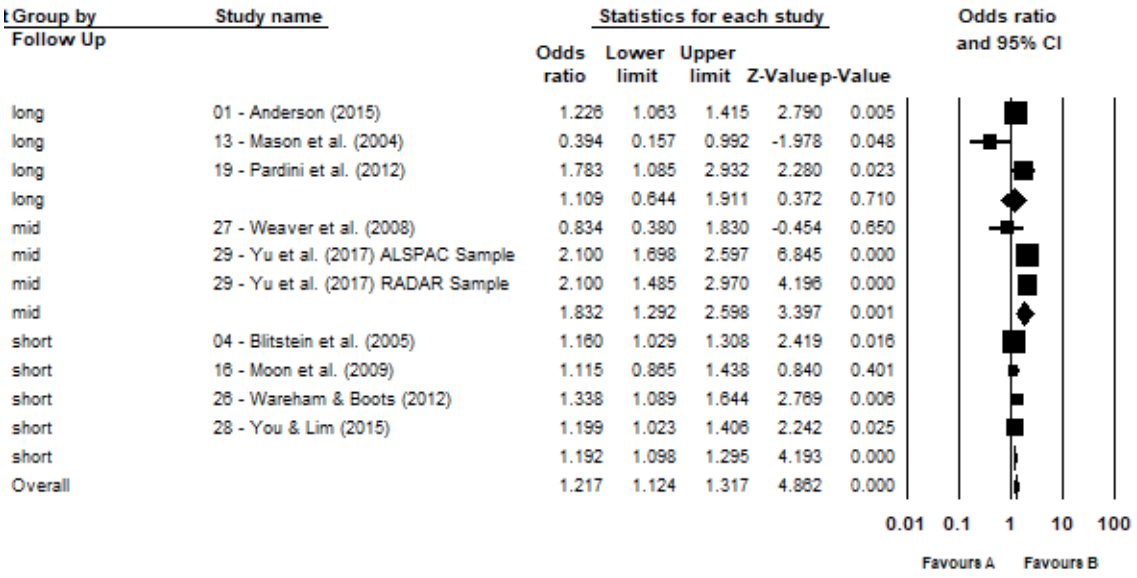

Figure C.13 Forest Plot: Odds of violent offending by follow up time length

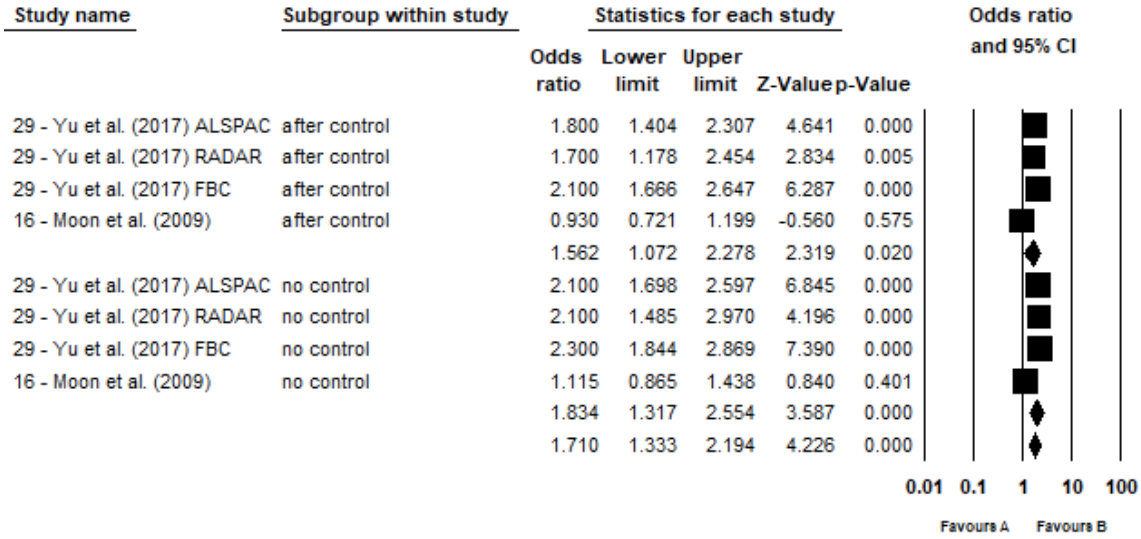

Figure C.14 Forest Plot: Odds of violent offending before and after controlling for confounding variables