CrossMark

SYSTEMATIC REVIEW

Protective Factors in Risk Assessment Schemes for Adolescents in Mental Health and Criminal Justice Populations: A Systematic Review and Meta-Analysis of their Predictive Efficacy

Geoffrey L. Dickens[1] · Laura E. O'Shea[1]

Abstract The consideration of protective factors has been integrated into a number of instruments whose aim is assess the risk of adverse outcomes among adolescents in high-risk mental health and criminal justice populations; however, little is known about their contribution to accurate risk prediction. We systematically reviewed the evidence for predictive efficacy of nine selected tools that require assessors to consider protective factors. Three tools had been tested for predictive ability but only one (the structured assessment of violence risk in youth) had been examined in multiple studies. Meta-analysis revealed that risk prediction based on the results did not improve over that based on a deficits model. Important decisions based on results of some protective factor-based tools should be treated with extreme caution since they lack empirical support. The importance of protective factors in problematic behavior has been demonstrated elsewhere, but this has not translated into significantly improved tools for use in clinical risk assessment in mental health and criminal justice populations.

Keywords Risk assessment · Protective factors · Violence · Self-harm · Meta-analysis · Systematic review

✉ Geoffrey L. Dickens
g.dickens@abertay.ac.uk

[1] School of Social and Health Sciences, Abertay University, Dundee DD1 1HG, UK

Introduction

Severe, problematic risk behaviors involving adolescents are a significant, international concern. Around 10% of adolescents, predominantly females, report having self-harmed (Hawton et al. 2002; Madge et al. 2008), and suicide, primarily by males, is the second most frequent cause of death in young people (Patton et al. 2009). Perpetration of violence peaks in late adolescence: the proportion of arrests for violent offences (homicide, rape, robbery, aggravated assault) in the US in 2015 who were under 18 years of age was 8.7%; 33.4% were under 25 years of age (United States Department of Justice 2017). The etiology of both self- and externally-directed aggression is complex. Nevertheless, psychiatric disorders, notably depression, anxiety, substance misuse, and Attention Deficit Hyperactivity Disorder, are common in young people presenting to general hospitals after self-harm; estimates of prevalence range from 48 to 87% (Al Ansari et al. 2001; Manor et al. 2010). In terms of youth violence, externalizing disorders including ADHD and oppositional defiant disorder (ODD) are known antecedents (Farrington 2005; Langbehn et al. 1998). In forensic and criminal/deviant youth populations, mental illness is especially common: two-thirds of males and three quarters of females in detention meet criteria for a clinical diagnosis (Teplin et al. 2006). A metaregression analysis of 25 surveys of youth aged 10–19 years in juvenile detention and correctional facilities found prevalence of psychotic illness to be 3.3 and 2.7% in boys and girls, respectively; 10.6 and 29.2% with major depression; and 11.7 and 18.5% with ADHD (Fazel et al. 2008). Higher rates of violent offending in people with psychotic disorders relative to siblings with no such disorder, and to matched general population controls, suggest an association between psychosis and violence. Further, higher rates
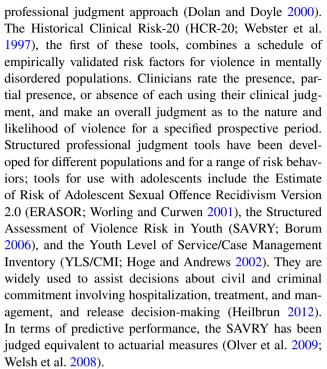
Springer

of violent offending among those with psychosis with the least time spent in hospital treatment implies a causal link. Against this backdrop, risk assessment for severe problematic behaviors is central to the work of professional mental health and criminal justice practitioners working with high-risk populations.

Risk assessment for the purpose of the prediction of adverse outcomes in forensic populations has its origins in the work of Ernest Burgess (1928). Burgess advocated an actuarial method of risk calculation in which individuals (adult parolees) were assigned a score of one for each of 21 characteristics (risk factors) that he deemed to be linked to parole violation; an overall risk level was assigned dependent on summed risk factors exceeding a predetermined threshold or "cut off" score. In a sample of 3000 parolees, 76% of those deemed high risk went on to re-offend in the subsequent 5-year period. In an era where individual clinical viewpoint ("unstructured clinical judgment") was the norm, Burgess' approach represented a step change inasmuch as it had the potential to inform decisions about parole. Glueck and Glueck (1950) used a similar method to examine risk factors in 500 delinquent boys and controls matched on important variables: age, ethnicity, intelligence, and income. Rather than simply assigning an equal score for the presence of each risk factor, weightings were derived from the strength of the risk factor as determined by the magnitude of difference between delinquents and controls. Thus, evidence of serious misbehavior in school (96% of delinquents versus 17% non-delinquents or 5.6:1) would be assigned more weight in the overall risk calculation than evidence of coming from a "broken home" (60 versus 34%; 1.8:1). The Gluecks viewed delinquency as essentially multifactorial in origin and approached the subject from a multidisciplinary perspective (Laud and Sampson 1991). Again, there are considerable advantages in terms of transparency and in the potential to improve decision-making. However, like other actuarial approaches, the risk factors selected tend to be static in nature (i.e., demographic features, prior convictions, and personality characteristics), they do not prioritize clinically relevant modifiable variables—sometimes termed dynamic factors—that could be targeted for treatment, and they reduce the role of professional judgment (Hart 1998a, b). Further, some view the assignment of risk levels or, more specifically, the subsequent restrictions on freedom that can result from them, based on the superficial similarity of an individual to others as unjust and therefore inherently unethical. If the ultimate aim of risk assessment is both to protect the public and promote rehabilitation by targeting modifiable risk factors for treatment then a reliance solely on actuarial methods might not be most effective (Silver and Miller 2002).

In the last 20 years, risk assessment has evolved to address these criticisms in the form of the structured professional judgment approach (Dolan and Doyle 2000). The Historical Clinical Risk-20 (HCR-20; Webster et al. 1997), the first of these tools, combines a schedule of empirically validated risk factors for violence in mentally disordered populations. Clinicians rate the presence, partial presence, or absence of each using their clinical judgment, and make an overall judgment as to the nature and likelihood of violence for a specified prospective period. Structured professional judgment tools have been developed for different populations and for a range of risk behaviors; tools for use with adolescents include the Estimate of Risk of Adolescent Sexual Offence Recidivism Version 2.0 (ERASOR; Worling and Curwen 2001), the Structured Assessment of Violence Risk in Youth (SAVRY; Borum 2006), and the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge and Andrews 2002). They are widely used to assist decisions about civil and criminal commitment involving hospitalization, treatment, and management, and release decision-making (Heilbrun 2012). In terms of predictive performance, the SAVRY has been judged equivalent to actuarial measures (Olver et al. 2009; Welsh et al. 2008).

One further development in risk assessment instrumentation for use with high risk populations has been the identification of the importance, and the subsequent inclusion, of protective factors. In contrast to the adult field, where the inclusion of protective factors in risk assessment for mental health and criminal justice populations is relatively new (de Vogel et al. 2011; Webster et al. 2009), their importance in adolescent risk assessment is longer standing (Arthur et al. 2003; Borowsky et al. 1997; Deković 1999; Loeber and Farrington 1998; Rolf 1985; Rosenberg 1987). The drivers of this development include the vast literature on adolescent resilience, the concerns of mental health and criminal justice youth-professionals about the lifelong consequences of early stigma, and the belief that young people may be more amenable to treatment and change than adults (Viljoen et al. 2012).

Conceptually, protective factors have been described in two main ways. First, defined simply as an absence of risk, they are viewed as individual and social characteristics that might be preventative by halting the occurrence of the risk factor(s) responsible for problem behavior. For example, excellent performance at school may be viewed as a protective factor against delinquency because it is the opposite of poor performance which is a risk factor (Shader 2001). Alternatively, protective factors have been viewed as characteristics that may reduce dysfunction directly or act as a buffer to mediate the negative effects of risk factors (Dignam and West 1988; Wheaton 1986). Further, protective factors may directly promote pro-social behaviors that, in themselves, may be protective (Jessor and Turbin 2014). The most comprehensive account thus far in the literature is

that of Jessor and Turbin (2014; see also Jessor et al. 1991; Jessor 2014) who describe the role of protective factors in the context of problem behavior theory (PBT) in which they comprise characteristics related to presence of pro-social role models, informal social and personal controls, social support for pro-social behavior, and engagement in pro-social behavior. Risk factors, however, are related to presence of role models for problem behavior, opportunities to engage in problem behavior, vulnerabilities for engagement in problem behavior, and actual engagement in problem behavior. Both risk and protective factors might emanate from family, peer, civic, and school environments. Protective factors are viewed not as simply absence of risk, but as distinct entities which can both influence problem- and prosocial-behaviors, and mitigate risk factors. Application of the theory to study of problem and prosocial behaviors in large US and Chinese samples of adolescents (Jessor and Turbin 2014) suggests that presence of informal social and personal controls is the best predictor of problem behavior; the presence of prosocial models and of social support predict prosocial behavior but not problem behavior. Thus, it seems probable that protective and risk factors may play somewhat different roles in the causation of both prosocial and problem behaviors.

Current Study

Despite the use of protective factors in theory and research, Jessor and Turbin (2014) have identified considerable variation and ambiguity in how protective factors have been conceptualized and minzed in risk assessment research studies. While protective factors have been studied in relation to a wide range of undesirable behaviors in adolescence, including tobacco use, sedentariness, truancy, and unhealthy eating (Jessor 2014), the current study is focused on their use in studies of the most severe and acutely risky behaviors such as self-harm and violence. The use of tools or schemes, i.e., schedules comprising empirically-derived risk factors with ratings guidance designed for use by mental health and criminal justice professionals conducting risk assessment of juveniles, has grown significantly in recent years: for example, more than 90% of US states now employ such tools in violence assessment (National Center for Juvenile Justice 2012). We are aware of a number of proprietary risk assessment schemes that are intended to guide in-depth assessment of adolescent mental health and criminal justice populations in regard to these behaviors, and that claim to incorporate protective factors in that assessment. As a first step in an attempt to consolidate current progress in the field, and to identify future research and development priorities, we have conducted a study to determine the extent to which these tools improve the predictive efficacy of assessment relative to that based on risk factors alone.

Method

Review Protocol

The review was conducted in accordance with the preferred reporting items for systematic reviews and meta-analyses statement (Moher et al. 2009) in order to facilitate transparent reporting. Included studies were selected as part of a larger literature search regarding the incorporation of protective factors in risk assessments in both adult and adolescent populations; a meta-analysis of the role of protective factors in adults has been reported on previously (O'Shea and Dickens 2016).

Tool Selection

We conducted extensive literature database and internet searching to identify tools that have been developed to assist with assessment of risk, and which explicitly include protective factors assessment. We identified 17 instruments that aim to assist mental health professionals in the assessment of protective factors. For the purpose of the current review, the identified tools designed for use with adolescent populations specifically ($n = 9$) were the short-term assessment of risk and treatability: adolescent version (START:AV; Viljoen et al. 2012), the structured assessment of violence risk in youth (SAVRY; Borum 2006), the San Diego regional resiliency check-up (SDRRC; Turner and Fain 2006), the multiplex empirically guided inventory of ecological aggregates for assessing sexually abusive children and adolescents (MEGA♪; Miccio-Fonseca 2013), the Columbia Suicide Severity Rating Scale (CSSRS; Posner et al. 2011), the reasons for living inventory-adolescents (RFL-A; Gutierrez et al. 2000), the Brief RFL-A (BRFL-Osman et al. 1996), the RFL-young adults (RFL-YA; Gutierrez et al. 2002), and the RFL-college students (RFL-CS; Westefeld et al. 1992).

Search Strategy

Multiple electronic databases (PsycINFO, Scopus, Web of Knowledge, Cochrane Library, CINAHL, and NCJRS) were searched for articles published before June 25 2014 as part of the wider search strategy. Search terms relating to the selected assessment tools were combined with those pertaining to multiple adverse outcomes (see example in online Appendix A). Wild card search terms (those ending with "*") were used to return all permutations of each search term. All articles, including "gray" literature

(e.g., conference presentations, technical reports, theses and dissertations), were eligible for inclusion and additional studies were located through hand searching reference lists of papers identified by the previous step.

Inclusion and Exclusion Criteria

Eligibility of articles was assessed by the second author as part of the previous review (O'Shea and Dickens 2016); the first author independently reviewed 25% of the studies to establish inter-rater reliability ($\kappa = 0.91$). Articles must have documented an original empirical investigation of the predictive validity of one or more of the identified tools for any of its intended outcomes in an adolescent population, using a prospective or pseudo-prospective design. Area under the curve (AUC) values and their associated 95% confidence intervals (CI) must have been included, or there must have been sufficient statistical information to allow for their calculation. Studies were excluded if they were not written in English, or if the assessment instruments had been amended or adapted from the published version. In cases where samples overlapped, only the study with the largest sample size was retained to avoid including the same participants twice; the exception to this were cases where different tools or outcomes were examined, in which case both studies were retained.

Operationalization of Protective Factors at Tool Level

For each tool that was used in at least one study of predictive ability we identified how protective factors were operationalized. We assigned tools to one of two categories used in a previous study (O'Shea and Dickens 2016): (A) Tool comprises factors that define a protective factor as lying at the opposite end of a continuum to a risk factor (e.g., history of kindness as opposed to history of violence measured on a single continuum); (B) Tool comprises protective factors which are defined as conceptually distinct from risk factors (e.g., history of kindness irrespective of history of violence). A third possibility, i.e., that protective factors are defined simply as the absence of a risk factor, was considered but rejected since this would encompass all formal risk assessment tools whose scoring is predicated on absence or presence of risk factors irrespective of any formal acknowledgement of protective factors. In addition, we identified how raters are guided to integrate protective factors into the overall risk prediction (e.g., through structured professional judgment into a risk estimate; or actuarial determination through cut-off scores).

Data Extraction for Meta-Analysis

The following information was extracted from included studies: number of participants, country of data collection, setting, length of follow-up period, assessment tool(s) used, adverse outcome(s) measured, sample characteristics, the AUC value and 95% CI for each protective scale-outcome combination. The AUC values and 95% CIs were also extracted for summary judgments and risk scales of tools that assess protective factors, in order to provide contextual information regarding the magnitude of effect sizes for protective factors.

Risk of Bias

The quality of included studies was independently assessed by both authors (linear weighted kappa = 0.95) using criteria developed by Haney et al. (2012). Each domain is rated as "yes", "unclear", or "no", and overall risk of bias (low, unclear, or high) reflects raters' opinions that identified biases reduce confidence of results (see Table 1).

Data Synthesis

Most studies either inverted protective scale scores or used these scales to predict desistance from risk behavior; for the remaining studies that did not use either of these procedures, we inverted AUC values to facilitate synthesis of results and comparisons with effect sizes from risk scales and, if relevant, summary judgments. Effect size was extracted for each scale-outcome combination. Given that outcomes were defined by the authors of each individual study, there may be some variation in definitions used. In order to minimize the number of reported outcomes the following effect sizes were combined: (1) inpatient physical aggression against others, inpatient physical aggression against objects, and inpatient physical aggression as "any inpatient aggression"; (2) inpatient verbal abuse and inpatient verbal threats as "inpatient verbal aggression"; and (3) non-sexual reoffending, non-violent reoffending, and general reoffending as "general reoffending". Where a study reported on more than one of the outcomes in any group, a mean effect size was calculated.

Meta-analysis was conducted using R (R Core Team 2015). AUC values were converted to Cohen's *d* values using a table provided by Rice and Harris (2005), which are a commonly used measure of effect size appropriate for random effects models (Yang et al. 2010). However, as Cohen's *d* values can be biased and overestimate the true effect, especially for small sample sizes (Lakens 2013), they were converted to Hedge's *g* values, which corrects for bias, using formulas provided in the compute.es package (Del Re 2013). Hedge's *g* values, weighted by the inverse

Table 1 Quality assessment of included studies

References	1. Adequate description of population	2. Non-biased selection	3. Low loss to follow-up/missing data	4a. Standardized method of risk factor assessment and scoring clearly described of referenced	4b. Unbiased risk factor assessment by independent assessors?	5a. Adequate outcome measurement?	5b. Unbiased outcome measurement by independent assessors?	6. Adequate accounting for potential confounders	Overall Assessment of potential for bias (low/unclear/high)
Adamson (2010)	No—does not report gender, ethnicity or age separately by group	Yes—all eligible	Yes—100% maintained	Yes	Yes—routine clinical practice	Yes	Unclear—recorded by clinical team but checked by psychologists who may have knowledge of risk assessment	No—no description of matching or controlling for differences	Unclear
Dolan and Rennie (2008)	Yes	Unclear—no description of consecutive or random sampling	Yes—99/123	Yes	Yes—blind to outcome data	Yes	Yes—blind to risk assessment	Yes—prospective assessment of single cohort	Unclear
Gammelgård et al. (2008)	Yes	Yes—all eligible/consecutive admissions	Yes—100% maintained	Yes	Unclear—no description of assessor independence or blinding	Yes	Unclear—no description of assessor independence or blinding	Yes—pseudo-prospective assessment of single cohort	Unclear
Hilterman et al. (2014)	Yes	Yes—random sampling	No—105/145 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—prospective assessment of single cohort	Unclear
Lodewijks et al. (2008a)	Yes	Unclear—no description of consecutive or random sampling	Yes—100% maintained	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	No—no description of matching or controlling for differences	Unclear
Lodewijks et al. (2008b)	Yes	Unclear—no description of consecutive or random sampling	Yes—117/130 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—pseudo-prospective assessment of single cohort	Unclear
Lodewijks et al. (2008c)	Yes	Yes—consecutive admissions	Yes—66/70 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	No	Unclear—no description of assessor independence or blinding	Yes—prospective assessment of single cohort	Unclear

Table 1 (continued)

References	1. Adequate description of population	2. Non-biased selection	3. Low loss to follow-up/missing data	4a. Standardized method of risk factor assessment and scoring clearly described of referenced	4b. Unbiased risk factor assessment by independent assessors?	5a. Adequate outcome measurement?	5b. Unbiased outcome measurement by independent assessors?	6. Adequate accounting for potential confounders	Overall Assessment of potential for bias (low/unclear/high)
Lodewijks et al. (2010)	Yes	Unclear—no description of consecutive or random sampling	Yes—100% maintained	Yes	Unclear—no description of assessor independence or blinding	Yes	Yes	No—no description of matching or controlling for differences	Unclear
McEachran (2001)	No—does not report gender, ethnicity or age for those with follow-up data	Yes—random sampling	No	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	No—no description of matching or controlling for differences	Unclear
McGowan et al. (2011)	Yes	Yes—all eligible	Yes—87/89 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	No—same researcher that coded risk assessment	Yes—prospective assessment of single cohort	Unclear
Meyers and Schmidt (2008)	Yes	Unclear—no description of consecutive or random sampling	Yes—121/133 had risk assessment and outcome data	Yes	Unclear—no description of assessor independence or blinding	Yes	Yes	Yes—pseudo-prospective assessment of single cohort	Unclear
Miccio-Fonseca (2013)	Yes	Unclear—no description of consecutive or random sampling	No—only 334/969 had follow-up data	Yes	Yes—blind to outcome data at point of coding	No	Yes	Yes—prospective assessment of single cohort	High
Owens (2011)	Yes	Unclear—no description of consecutive or random sampling	Yes—93/100 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—pseudo-prospective assessment of single cohort	Unclear
Penney et al. (2010)	No—does not report age etc. by group	Unclear—no description of consecutive or random sampling	Yes—100% maintained	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	No—no description of matching or controlling for differences	Unclear
Rennie and Dolan (2010)	Yes	Unclear—no description of consecutive or random sampling	Yes	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—prospective assessment of single cohort	Unclear

Table 1 (continued)

References	1. Adequate description of population	2. Non-biased selection	3. Low loss to follow-up/missing data	4a. Standardized method of risk factor assessment and scoring clearly described of referenced	4b. Unbiased risk factor assessment by independent assessors?	5a. Adequate outcome measurement?	5b. Unbiased outcome measurement by independent assessors?	6. Adequate accounting for potential confounders	Overall Assessment of potential for bias (low/unclear/high)
Richard (2013)	Yes	No—recruited through poster advertising	Yes—235/241 had risk assessment and outcome data	Yes	Unclear—no description of assessor independence or blinding	Yes	Yes	Yes—pseudo-prospective assessment of single cohort	Unclear
Schmidt et al. (2011)	No—does not report age etc. by group	Yes—consecutive referrals	Yes—minimum of 112/133 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—prospective assessment of single cohort	Low
Shepherd et al. (2014a)	Yes	Unclear—no description of consecutive or random sampling	Yes—175/177 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—prospective assessment of single cohort	Unclear
Shepherd et al. (2014b)	Yes	Unclear—no description of consecutive or random sampling	Yes—213/215 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—prospective assessment of single cohort	Unclear
Viljoen et al. (2008)	Yes	Yes—consecutive admissions	Yes—100% maintained	Yes	Unclear—no description of assessor independence or blinding	Yes for offense data, unclear for violence during treatment	Yes for offense data, unclear for violence during treatment	Yes—pseudo-prospective assessment of single cohort	Unclear
Viljoen et al. (2012)	No—does not report gender, ethnicity or age for those with follow-up data	Unclear—no description of consecutive or random sampling	Yes—81/90 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes for all outcomes	Yes for offending Unclear for remaining outcomes	Yes—prospective assessment of single cohort	Unclear
Vincent et al. (2012)	No—does not report mean ages	Yes—all eligible	Yes—674/712 had risk assessment and outcome data	Yes	Yes—routine practice	Yes	Yes	Yes—prospective assessment of single cohort	Low
Welsh et al. (2008)	Yes	Yes—consecutive referrals	Yes—105/133 had risk assessment and outcome data	Yes	Yes—blind to outcome data at point of coding	Yes	Yes	Yes—pseudo-prospective assessment of single cohort	Low

variance weight, were pooled using the rma.mv function from the metafor package (Viechtbauer 2010). This conducts a multilevel meta-analysis that can take into account non-independence of effect sizes resulting from the nested data structure in the current review. Random effects were included for outcome and scale nested within study since estimates derived from the same scale, or predicting the same outcome, will likely be correlated. Random effects were also added for study nested within author, and for author, since effect sizes within studies and authors are likely to be more similar than those between studies or authors. Effect sizes were estimated for each outcome relative to that for any inpatient aggression because the ability of structured professional judgment schemes for prediction of this outcome is well established (e.g., Singh et al. 2011). Effect sizes for protective scales and summary judgments were estimated with risk scales as the reference category because their relative performance is of interest. The effect of gender, coded as the percentage of the sample that is female, was included as a moderator as previous research has found that some risk assessment schemes perform more accurately in women (O'Shea et al. 2013; O'Shea and Dickens 2015). The risk of bias as determined by quality assessment was also included as a moderator to investigate the effect of bias on estimated effect size. Effect sizes were classified as small (0.2), moderate (0.5), and large (0.8) according to Cohen's criteria (Cohen 1992). Equivalent AUC values were presented for estimated effect sizes to facilitate comparison with previous research and I^2 values were calculated to quantify the extent of any observed heterogeneity (Heudo-Medina et al. 2006).
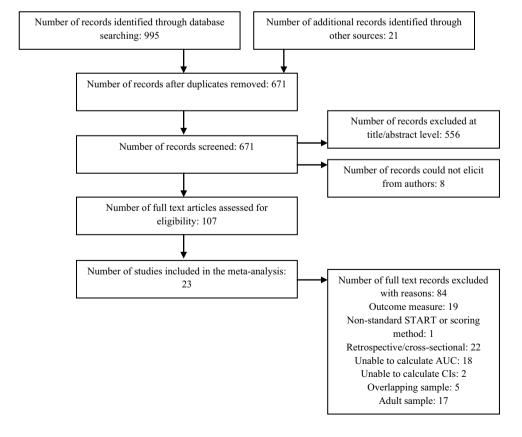
Results

Study Characteristics

The literature search conducted for the previous review (O'Shea and Dickens 2016) identified 107 articles for which the full-text was reviewed; 67 records were excluded resulting in 17 studies for inclusion in the meta-analysis of the contribution of protective factors to risk assessment in adults (O'Shea and Dickens 2016), and 23 studies, comprising 30 independent samples, for inclusion in the current review (see Fig. 1 for exclusion reasons).

The total sample size was 3280 (mean $N = 143$); four articles were masters or doctoral theses and 19 were journal articles published between 2008 and 2014. The most researched tool was the SAVRY ($k = 21$); the START:AV and MEGA♪ were both examined by only one study. None of the included studies examined the predictive validity of the SDRRC, CSSRS or any of the RFL variants. Studies were conducted in Canada ($k = 9$), the United Kingdom ($k = 4$), the United States of America ($k = 4$), Australia

Fig. 1 Flow diagram of literature search: Modified from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement flow diagram (Moher et al. 2009). AUC, area under the receiver operating characteristic curve

Number of records identified through database searching: 995

Number of additional records identified through other sources: 21

Number of records after duplicates removed: 671

Number of records screened: 671

Number of records excluded at title/abstract level: 556

Number of records could not elicit from authors: 8

Number of full text articles assessed for eligibility: 107

Number of studies included in the meta-analysis: 23

Number of full text records excluded with reasons: 84
Outcome measure: 19
Non-standard START or scoring method: 1
Retrospective/cross-sectional: 22
Unable to calculate AUC: 18
Unable to calculate CIs: 2
Overlapping sample: 5
Adult sample: 17

($k = 2$), Finland ($k = 1$), and Spain ($k = 1$). Samples were drawn from juvenile justice facilities ($k = 10$), psychiatric and forensic psychiatric inpatient units ($k = 7$), probation services ($k = 5$), custody ($k = 4$), assessment centres ($k = 3$), correctional or specialist schools ($k = 2$), and an outpatient psychology clinic ($k = 1$). The most commonly studied outcomes were "any violent offending" ($k = 14$ studies) and "general reoffending" ($k = 10$). Eleven other outcomes related to inpatient behavior, sexual recidivism, technical recidivism, self-harm, victimization, substance/street drug use were examined in between one and three studies each. See Table 2 for full study characteristics.

Characteristics of Included Tools

All three tools operationalize protective factors as distinct from risk factors (definition A). SAVRY and START:AV recommend incorporation of identified protective factors into a risk formulation through a process of structured professional judgment. In the MEGA♪, protective factors are integrated into a risk formulation through actuarial methods.

The SAVRY (Borum 2006)

The SAVRY comprises 30 items; 10 historical (e.g., history of violence, childhood history of maltreatment), six social (e.g., peer delinquency, poor parental management), eight individual/clinical (e.g., negative attitudes, anger management problems), and six protective (prosocial involvement, strong social support, strong attachments and bonds, positive attitude toward intervention and authority, strong commitment to school, resilient personality traits). It is recommended for use in adolescents aged between 12 and 18 years to assess risk of violence. Each risk item is rated as low (0), moderate (1), or high (2); the protective items are scored as absent (0) or present (1). For research purposes scale scores can be calculated by summing individual item scores; a total risk score can also be calculated by summing the historical, social, and clinical scales. Two studies (Lodewijks et al. 2010; Vincent et al. 2012) also calculated a dynamic items score by summing the social and individual scale items. Finally assessors are required to make an overall rating of low, moderate, or high risk for future violence based on presence and relevance of items and a range of additional information; one study (Hilterman et al. 2014) also reported a risk estimate of low, moderate, or high risk of general offending.

The START:AV (Viljoen et al. 2012)

The START:AV is a structured professional judgment tool comprising 23 dynamic items (sample items: School and work, Recreation, Substance use, Rule adherence) that pertain to adolescents and their social contexts and is adapted from the adult START (Webster et al. 2009). Each item is scored on two 3-point scales, once in terms of protective factors (strengths) and once in terms of risk (vulnerabilities). Specific risk estimates (SREs; low, moderate, or high) are then formed from consideration of the presence and relevance of the 23-items plus a range of additional information regarding the likelihood of eight adverse outcomes occurring over a maximum of three months (violence, self-harm, suicide, self-neglect, substance abuse, unauthorized leave, and victimization, general offending). Scoring criteria were adapted from the START to reflect adolescents' developmental context and outcomes were adjusting to include risk behaviors relevant to adolescents, such as running away from home rather than the START-item "unauthorized leave".

MEGA♪ (Miccio-Fonseca 2013, 2016; Miccio-Fonseca and Rasmussen 2013, 2015)

MEGA♪ is a 75-item tool intended to aid assessment of risk for sexually inappropriate or abusive behaviors in children and adolescents aged 4–19 years. It comprises four scales: (1) the risk scale, containing 45 historical and dynamic items assessing generalized risk for "coarse sexual improprieties and/or sexually abusive behaviors" (p. 627); (2) a protective scale, comprising ten historical and dynamic items (e.g., youth is rule bound, getting along better with others) that mitigate risk; (3) an estrangement scale containing 14 items related to family relationships assessing whether the use is a victim of any type of abuse (e.g., parental separation, exposure to domestic violence); and (4) the 6-item persistent sexual deviancy scale which captures the frequency and progression of sexually abusive behaviors (e.g., offender-victim age disparity). It is recommended that MEGA♪ be completed every 6 months after a case file review and interview where possible.

Risk of Bias

Study quality assessment indicated that the vast majority ($k = 19$) of studies were rated as unclear risk of bias; three studies were rated as low risk of bias and one was rated as high risk (see Appendix A). The most common source of potential bias was failure to provide evidence that participants had been randomly selected or were consecutive admissions.

Individual Study Effect Sizes

A total of 278 individual AUC values were contributed from the 23 studies. The number of AUC values

Table 2 Characteristics of studies included in meta-analysis

References	Risk instrument	N	Country	Population Characteristics	Setting	Length of follow-up	Outcomes predicted
Adamson (2010)	SAVRY[a,b,c,d,e]	30	UK	Intellectual disability (ID) and autistic spectrum disorders (ASD)	Forensic secure psychiatric	3 months	Any inpatient aggression
Dolan and Rennie (2008)	SAVRY[a,b,c,d,e,f]	99	UK	Mean age 16.15; 100% male; 83.8% Caucasian	Custody	12 months	Any violent offending; General reoffending
Gammelgård et al. (2008)	SAVRY[a,b,c,d]	208	Finland	Mean age: 15.2; 57.7% male	Secure psychiatric, correctional schools and forensic psychiatric	6 months	Any inpatient aggression
Hilterman et al. (2014)	SAVRY[a,b,c,d,e,f,g]	105	Spain	Mean age: 18.4; 82% male	Probation	12 months	Any violent offending; general reoffending
Lodewijks et al. (2008a)	SAVRY[a,b,c,d,e,f]	35; 47	Netherlands	Mean age: 17.2, 17.6; 0% male, 100% male; 40% conduct disorder, 40% oppositional defiant disorder, 60% other Axis I disorder; 38% conduct disorder, 49% ODD, 53% other Axis I disorder; 66% Caucasian, 57% Caucasian	Juvenile justice	Mean 18 months, Mean 16.6 months	Any violent offending
Lodewijks et al. (2008b)	SAVRY[a,b,c,d,e,f]	117	Netherlands	Mean age: 15.3; 95% male; 28% Caucasian	Juvenile justice	36 months	Any violent offending
Lodewijks et al. (2008c)	SAVRY[a,b,c,d,e]	66	Netherlands	Mean age: 15.4; 100% male; 62% Caucasian	Psychiatric inpatient	–	Inpatient verbal aggression
Lodewijks et al. (2010)	SAVRY[a,f,h]	42	Netherlands	Mean age: 17.5; 100% male; 65% Caucasian	Juvenile justice	Mean 16.5 months	Any violent offending
McEachran (2001)	SAVRY[a,b,d,e,f]	36	Canada		Inpatient assessment unit	36 months	Any violent offending
McGowan et al. (2011)	SAVRY[a,b,c,d,f]	87	USA	Mean age: 15.0; 82% male; 94.3% emotional disability, 3.1% ID, 1.1% ASD, 1.1% other health impairment; 71% Caucasian	Specialist school	12 months	Inpatient physical aggression
Meyers and Schmidt (2008)	SAVRY[c]	121	Canada	Mean age 14.9; 66% male, 69% Caucasian	Juvenile justice	12 months	Any violent offending; general reoffending
Miccio-Fonseca (2013)	MEGA♪[i]	334	USA, Canada, UK	Mean age: 14.5; 91.6% male; 53.6% Caucasian	Juvenile justice, psychiatric inpatient, outpatient psychology	6 months	Sexual recidivism
Owens (2011)	SAVRY[a,b,c,d,e,f]	93	USA	Mean age: 15.5; 53% Caucasian	Juvenile justice	75.6 months	General reoffending; sexual recidivism
Penney et al. (2010)	SAVRY[a,b,c,d,e]	80; 64	Canada		Custody, assessment centres and probation	2 years	Any violent offending
Rennie and Dolan (2010)	SAVRY[f]	111	UK	Mean age 16.14; 100% male; 84.4% Caucasian	Custody	12 months	General reoffending
Richard (2013)	SAVRY[a,b,c,d,f]	235	Canada	Mean age 17.1; 100% male; 76.1% Caucasian	Juvenile Justice	12 months	General reoffending

Table 2 (continued)

References	Risk instrument	N	Country	Population Characteristics	Setting	Length of follow-up	Outcomes predicted
Schmidt et al. (2011)	SAVRY[c,e,f]	80; 48	Canada		Assessment center	124.8 months	Any violent offending; general reoffending; Technical recidivism
Shepherd et al. (2014a)	SAVRY[a,b,c,d,e,f]	84; 59; 32	Australia	Mean age 20.0, 17.3, 16.3; 100% male	Juvenile justice	Minimum 6 months	Any violent offending; general reoffending
Shepherd (2014b)	SAVRY[a,b,c,d,e,f]	38	Australia	0% male	Juvenile justice	Minimum 6 months	Any violent offending; general reoffending
Viljoen et al. (2008)	SAVRY[a,b,c,d,e,f]	169	Canada	Mean age: 15.4; 100% male; 83.4% Caucasian	Psychiatric inpatient	Inpatient outcomes mean 12.8 months; Community outcomes mean 78.96 months	Any inpatient aggression; any violent offending; sexual recidivism; inpatient sexual aggression; any reoffending
Viljoen et al. (2012)	START:AV[j,k,l,m,n]	81	Canada		Probation	3 months	Self-harm in community; victimization in community; substance use in community; any reoffending; Street drug use in community; any violent offending
Vincent et al. (2012)	SAVRY[a,c,e,h]	38; 254; 382	USA	Age ranges <12, 13–15, 16–18; 65.8% male, 71.7% male, 74.3% male	Probation	Mean 14.5 months	General reoffending; any reoffending; technical recidivism; any violent offending
Welsh et al. (2008)	SAVRY[c]	105	Canada	Mean age: 14.6; 63.8% male; 71.4% Caucasian	Assessment center	Mean 35.8 months	General offending; any violent offending

[a]SAVRY historical scale total
[b]SAVRY individual scale total
[c]SAVRY risk scale total
[d]SAVRY social scale total
[e]SAVRY violence risk estimate
[f]SAVRY protective scale total
[g]SAVRY general risk estimate
[h]SAVRY dynamic items total
[i]MEGA♪ total
[j]START strengths scale total
[k]START substance use risk estimate
[l]START victimization risk estimate
[m]START violence risk estimate
[n]START vulnerability scale total

contributed ranged from 1 to 48. The magnitude of AUC values ranged from 0.44 to 0.91; 171 (62%) of these were significantly greater than chance.

Mean Weighted Effect Sizes

Results of the meta-analysis are presented in Table 3. Moderators accounted for a significant proportion of the heterogeneity of effect sizes ($Q[54] = 79.17$, $p = 0.014$; $I^2 = 33\%$) but there was still a significant amount of unexplained heterogeneity ($Q[223] = 276.25$, $p = 0.009$; $I^2 = 20\%$). Estimated Hedge's g value when values were equivalent to the reference categories (i.e., outcome = any inpatient aggression; scale type = risk; scale = MEGA♪ total; bias = low) was 1.38 (AUC = 0.84). There was no significant effect of scale type on estimated effect size ($Q[2] = 1.57$, $p = 0.457$;

$I^2 = 0\%$); however, effect sizes were smaller for protective scales and summary judgments compared with risk scales. There was a significant effect of individual scale ($Q[12] = 22.20$, $p = 0.035$; $I^2 = 50\%$) with the SAVRY historical (-0.53, $p = 0.017$) and individual (-0.46, $p = 0.038$) scales performing significantly poorer than the MEGA♪ total score. Overall, estimates of effect size were not moderated by outcome ($Q[12] = 19.77$, $p = 0.072$; $I^2 = 44\%$); however, estimated effect sizes were significantly poorer for a number of outcomes relative to any inpatient aggression: any reoffending (-0.44, $p = 0.004$), violent reoffending (-0.45, $p = 0.002$), general reoffending (-0.41, $p = 0.007$), sexual reoffending (-0.65, $p < 0.001$), inpatient physical aggression (-0.49, $p = 0.043$), inpatient sexual aggression (-0.54, $p = 0.006$), and inpatient verbal aggression (-1.00, $p = 0.04$). Estimates of effect size were not moderated by

Table 3 Estimated mean weighted effect sizes relative to reference category

	Estimate	SE	95% CI	p
Intercept[a]	1.38	0.38	[0.64, 2.12]	<0.001
Outcome				
Any reoffending	−0.44	0.15	[−0.75, −0.14]	0.004
Violent reoffending	−0.45	0.15	[−0.74, −0.16]	0.002
General reoffending	−0.41	0.15	[−0.71, −0.11]	0.007
Inpatient physical aggression	−0.49	0.24	[−0.96, −0.02]	0.043
Inpatient sexual aggression	−0.54	0.20	[−0.93, −0.15]	0.006
Inpatient verbal aggression	−1.00	0.34	[−1.67, −0.32]	0.004
Self-harm in community	−0.45	0.44	[−1.31, 0.42]	0.310
Sexual recidivism	−0.65	0.19	[−1.03, −0.27]	0.001
Street drug use community	−0.22	0.49	[−1.17, 0.74]	0.658
Substance use community	−0.60	0.46	[−1.5, 0.29]	0.187
Technical recidivism	−0.31	0.18	[−0.66, 0.04]	0.085
Victimization in community	−0.45	0.40	[−1.24, 0.34]	0.268
Scale type				
Summary judgment	−0.44	0.35	[−1.12, 0.25]	0.214
Protective	−0.09	0.51	[−1.09, 0.92]	0.868
Scale				
SAVRY dynamic	−0.40	0.23	[−0.84, 0.05]	0.080
SAVRY general risk estimate	−0.38	0.30	[−0.97, 0.21]	0.211
SAVRY historical	−0.53	0.22	[−0.96, −0.09]	0.017
SAVRY individual	−0.46	0.22	[−0.9, −0.03]	0.038
SAVRY protect	−0.68	0.44	[−1.53, 0.18]	0.124
SAVRY risk total	−0.41	0.22	[−0.85, 0.02]	0.061
SAVRY social	−0.71	0.44	[−1.58, 0.15]	0.106
SAVRY violent risk estimate	−0.39	0.24	[−0.87, 0.08]	0.103
START:AV self-harm RE	0.32	0.69	[−1.04, 1.67]	0.647
START:AV strength	−0.13	0.51	[−1.14, 0.88]	0.805
START:AV substance abuse RE	1.14	0.78	[−0.39, 2.67]	0.144
START:AV victimization RE	0.17	0.54	[−0.89, 1.22]	0.754

SE standard error, *95% CI* 95% confidence interval

[a]Estimated Hedge's g value for reference category (i.e. Outcome = violent offending, scale type = risk, bias = low risk)

gender ($Q[6] = 6.16$, $p = 0.405$; $I^2 = 19\%$), bias ($Q[1] = 0.19$, $p = 0.664$), or the interaction between scale type and outcome ($Q[20] = 12.89$, $p = 0.882$; $I^2 = 0\%$).

Estimated mean weighted Hedge's g values for each outcome are presented in Fig. 2. The largest effect size was for any inpatient aggression (1.32, AUC = 0.83) and the smallest was for inpatient verbal aggression (0.38, AUC = 0.61). All of the effect sizes were large, with the exception of those for inpatient verbal aggression, sexual reoffending, and substance use in the community; further, all estimates were significantly greater than 0, based on inspection of 95% confidence intervals, apart from inpatient verbal aggression and substance use in the community.

Discussion

Structured tools to assist with the prediction of severe outcomes including violence and self-harm are increasingly used by mental health and criminal justice professionals working with adolescents. Some of these schemes have explicitly integrated theories and concepts about protective factors into their construction and guidance documentation. The evidence about whether, and the extent to which, these tools improve predictive accuracy has not thus far been collated. We therefore aimed to synthesize the evidence for the predictive efficacy of structured risk assessment schemes, selected on the basis of their explicit intention of facilitating protective factors, for use in adolescent mental health and criminal justice populations. Our systematic search strategy revealed that there is currently no empirical evidence for the predictive efficacy of the SDRRC, CSSRS, RFL-A, Brief RFL-A, RFL-YA, or RFL-CS. The first conclusion we can draw therefore is that clinical decisions based on the use of these tools are not evidence-based; the

case for interventions or resource allocation predicated on them should be treated with extreme caution.

Each of the three tools actually subjected to one test or more of predictive efficacy (SAVRY, START-AV, MEGA♪) conceptualizes protective factors as distinct and separate from risk factors inasmuch as protective factors comprise a separate scale from each tools' risk factor scale(s). This is congruent with problem behavior theory (Jessor 2014), which describes risk and protective factors as direct, independent predictors of involvement in problem behavior and pro-social behavior; and protective factors as also moderating the influence of risk factors on involvement in these behaviors. However, the integration of the protective factors identified as relevant to the individual under assessment is handled differently across tools. Both the SAVRY and the START-AV are structured professional judgment tools whose guidance indicates that practitioners should consider all contributing elements in the formulation of a risk level. The MEGA♪ is based on actuarial principles with risk level assignment dictated from gender- and age-specific cutoff scores; the one included study of the MEGA♪ lacked some transparency about the contribution of the protective scale to the risk level assignment. For the START:AV and SAVRY, our results revealed that protective factor scales, and summary judgments based on a combined consideration of risk and protective factors, performed no better than the risk scales for the "any inpatient aggression" reference category.

Given that we know a considerable amount about the role of protective factors for a range of problem behaviors or conditions including depression (Cairns and Yap 2014), substance abuse (Beyers et al. 2004), and internet addiction (Koo and Kwon 2014) for the wider adolescent population, it is disappointing that their integration into assessment schemes for high risk populations, and for some very severe and acute outcomes, has not
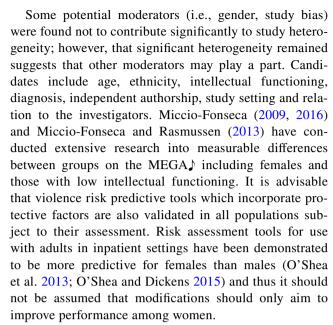
Fig. 2 Forest plot of estimated effect sizes by outcome

Outcome	AUC	Hedge's g [95%CI]
Any inpatient aggression	0.83	1.38 [0.64, 2.12]
Any reoffending	0.75	0.94 [0.64, 1.24]
Violent reoffending	0.74	0.92 [0.63, 1.21]
General reoffending	0.75	0.97 [0.67, 1.27]
Inpatient physical aggression	0.74	0.89 [0.42, 1.36]
Inpatient sexual aggression	0.72	0.84 [0.45, 1.23]
Inpatient verbal aggression	0.61	0.38 [-0.30, 1.06]
Self-harm in community	0.75	0.93 [0.07, 1.79]
Sexual reoffending	0.7	0.73 [0.35, 1.11]
Street drug use in community	0.79	1.16 [0.20, 2.12]
Substance use in community	0.71	0.77 [-0.13, 1.67]
Technical recidivism	0.77	1.07 [0.72, 1.42]
Victimisation in community	0.75	0.93 [0.14, 1.72]

demonstrably led to improved prediction. This is consistent with findings from our previous review of protective factor-oriented instruments for adult mental health and criminal justice populations (O'Shea and Dickens 2016). This might suggest that risk assessment schemes for both adolescents and adults in high-risk populations need to better integrate protective factor-related theory with risk assessment as a first step in an attempt to facilitate more accurate assessment. Extra urgency can be implied from findings that risk assessment tools with no explicit consideration of protective factors, while reasonably good at identifying low risk individuals, are limited in their ability to accurately identify those at high risk of, for example, violence (Fazel et al. 2012); better integration of protective factors might be one way of improving this situation. We propose that this requires new research in a number of key areas, and, if appropriate, translation into new protective factor-oriented assessment schemes. Notable questions for investigation are: (1) are the most relevant protective factors for the most relevant outcomes in high risk youth populations identified; (2) are they operationalized in a manner congruent with relevant theory; (3) is current theory sufficient to explain the most severe/acute outcomes in these high risk groups; (4) is the operationalization and identification communicated to assessors such that their practice is, in practice, fidelitous to the conceptualization; and (5) do new tools improve prediction?

The current study design did not allow us to investigate whether protective scales contribute uniquely to the formulation of a more accurate SRE, although this seems unlikely given the failure of SREs to outperform risk scales. While we have previously found evidence (O'Shea et al. 2013) for statistically significant incremental validity of the Strength scale of the original adult Short-Term Assessment of Risk and Treatability (START; Webster et al. 2009) in the case for prediction of violence, this was marginal, leading only to improved classification of violent and non-violent forensic inpatients of 1.6%. The tools examined in the current review appear to be conceptually premised on an empirically supported model of risk and protective factors (Jessor 2014). However, what appears to be in question is the manner in which, in these specific tools, those risk and protective factors are considered in formulation of a risk estimate given that this is conducted using structured professional judgment, a somewhat opaquely described method in research studies, and one which actively promotes rater discretion and use of clinical expertise (e.g., Guy et al. 2012). One of the major consequences of our study, therefore, is a fresh spotlight on the need to examine and articulate how decisions are made in structured professional judgment approaches including how risk and protective factors are weighed and synthesized into an overall risk estimate.

Some potential moderators (i.e., gender, study bias) were found not to contribute significantly to study heterogeneity; however, that significant heterogeneity remained suggests that other moderators may play a part. Candidates include age, ethnicity, intellectual functioning, diagnosis, independent authorship, study setting and relation to the investigators. Miccio-Fonseca (2009, 2016) and Miccio-Fonseca and Rasmussen (2013) have conducted extensive research into measurable differences between groups on the MEGA♪ including females and those with low intellectual functioning. It is advisable that violence risk predictive tools which incorporate protective factors are also validated in all populations subject to their assessment. Risk assessment tools for use with adults in inpatient settings have been demonstrated to be more predictive for females than males (O'Shea et al. 2013; O'Shea and Dickens 2015) and thus it should not be assumed that modifications should only aim to improve performance among women.

The studied tools examined a range of outcomes and predictive efficacy for outcomes significantly different from the reference category "inpatient physical aggression" (i.e., inpatient verbal aggression, sexual reoffending, and substance use in the community) were significantly poorer. It is somewhat unlikely that items included in the risk assessment schemes are equally predictive of all the outcomes they intend to cover. For the adult version of the START it has been demonstrated that subsets of items are significantly predictive of different outcomes (Braithwaite et al. 2010). Tool developers may wish to consider whether tools require modification and further validation for such outcomes; alternatively, tools should be very specific about which outcomes they have predictive validity for.

There were a number of records that we could not obtain, despite attempted communication with the authors; however, we did review a number of articles that would be considered as grey literature and ultimately included four theses, reducing the likelihood of publication bias. The overall quality of included studies was disappointing, with 20 of the 23 studies being rated as unclear or high risk of bias. Further, there were a number of scale-outcome combinations where there was only one available study which limited our ability to conduct more detailed analysis examining the interaction between outcome and scale. However, these criticisms are directed more at the state of the existing literature than at the current review per se, and highlight the need for more high quality studies investigating the role of protective factors for adverse outcomes in vulnerable adolescents.

Conclusion

The current study demonstrates that the integration of the assessment of protective factors in risk assessment among adolescents under the care of mental health and criminal justice services is seriously underdeveloped compared with studies of the predictive value of violence and sexual recidivism risk assessment tools (e.g., Olver et al. 2009). However, the importance of protective factors is more well-established in adolescent populations than adult ones, and there is good empirical support for at least one theoretical model (Jessor and Turbin 2014). In summary, the operationalization and predictive efficacy of protective factors in high risk mental health populations is currently of extremely limited value. There is considerable work to do to better integrate the wider protective factors research into relevant risk assessment tools in order to ensure resources are used appropriately and risk management is not disproportionately restrictive.

Author contributions   GLD conceived of the study, participated in its design and coordination and drafted the manuscript; LO conceived of the study, participated in the study design, performed the statistical analysis, and helped to draft the manuscript. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest   The authors report no conflicts of interest.

References

Adamson, L. G. (2010). *Investigating the assessment and treatment of violence in adolescents with developmental disabilities*. University of Birmingham. http://etheses.bham.ac.uk/728/10/Adamson-10PhD_redacted.pdf. Accessed 27 Mar 2017.

Al Ansari, A. M., Hamadeh, R. R., Matar, A. M., Marhoon, H., Buzaboon, B. Y., & Raees, A. G. (2001). Risk factors associated with overdose among Bahraini youth. *Suicide and Life Threatening Behavior, 31*, 197–206.

Arthur, M. W., Ayers, C. D., Graham, K. A., Hawkins, D. J., Bukoski, W. J., & Slobada, Z. (2003). *Mobilizing communities to reduce risks for drug abuse: a comparison of two strategies. Handbook of drug abuse prevention: Theory, science and practice* (pp. 129–144). New York: Kluwer Academic (**Plenum Publishers**).

Beyers, J. M., Toumbourou, J. W., Catalano, R. F., Arthur, M. W., & Hawkins, J. D. (2004). A cross-national comparison of risk and protective factors for adolescent substance use: the United States and Australia. *Journal of Adolescent Health, 35*, 3–16. doi:10.1016/j.jadohealth.2003.08.015.

Borowsky, I. W., Hogan, M., & Ireland, M. (1997). Adolescent sexual aggression: Risk and protective factors. *Pediatrics*, *100*(6), e7. doi:10.1542/peds.100.6.e7.

Borum, R. (2006). *Manual for the structured assessment of violence risk in youth (SAVRY)*. Odessa, Fl: Psychological Assessment Resources.

Braithwaite, E., Charette, Y., Crocker, A. G., & Reyes, A. (2010). The predictive validity of clinical ratings of the short term assesssment of risk and treatability (START). *International Journal of Forensic Mental Health, 9*(4), 271–281. doi:10.1080/14999013.2010.534378.

Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce (Ed.), *The workings of the indeterminate sentence law and parole in Illinois* (pp. 205–249). Springfield: Illinois State Parole Board.

Cairns, K. E., & Yap, M. B. H. (2014). Risk and protective factors for depression that adolescents can modify: A systematic review and meta-analysis of longitudinal studies. *Journal of Affective Disorders, 169*, 61–75. doi:10.1016/j.jad.2014.08.006.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155.

de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y.H.A. (2011). Assessing protective factors in forensic psychiatric practice. Introducing the SAPROF. *International Journal of Forensic Mental Health, 10*, 171–177. doi:10.1080/14999013.2011.600230.

Deković, M. (1999). Risk and protective factors in the development of problem behavior during adolescence. *Journal of Youth and Adolescence, 28*, 667–685. doi:10.1023/A:1021635516758.

Del Re, A. C. (2013). compute.es: Compute Effect Sizes. R package version 0.2-2. http://cran.r-project.org/web/packages/compute.es.

Dignam, J. T., & West, S. G. (1988). Social support in the workplace: Tests of six theoretical models. *American Journal of Community Psychology, 16*, 701–724.

Dolan, M., & Doyle, M. (2000). Violence risk prediction: Clinical and actuarial measures and the role of the Psychopathy Checklist. *British Journal of Psychiatry, 177*, 303–311. doi:10.1192/bjp.177.4.303.

Dolan, M. C., & Rennie, C. E. (2008). The structured assessment of violence risk in youth as a predictor of recidivism in a United Kingdom cohort of adolescent offenders with conduct disorder. *Psychological Assessment, 20*(1), 35–46. doi:10.1037/1040-3590.20.1.35.

Farrington, D. (2005). *Integrated developmental and life-course theories of offending: Advances in criminological theory (Vol. 14)*. New Brunswick, NJ: Transaction.

Fazel, S., Doll, H., & Langstrom, N. (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: A systematic review and metaregression analysis of 25 surveys. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 1010–1019. doi:10.1097/CHI.ObO13e31817eecf3.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behavior in 73 samples involving 24 827 people: systematic review and meta-analysis. *BMJ (Clinical Research Ed.), 345*, e4692. doi:10.1136/bmj.e4692.

Gammelgård, M., Koivisto, A. M., Eronen, M., & Kaltiala-Heino, R. (2008). The predictive validity of the structured assessment of violence risk in youth (SAVRY) among institutionalised adolescents. *Journal of Forensic Psychiatry and Psychology, 19*(3), 352–370. doi:10.1177/0093854815616842.

Glueck, S., & Glueck, E. (1950). *Unravelling juvenile delinquency*. Cambridge, MA: Harvard University Press.

Gutierrez, P. M., Osman, A., Barrios, F. X., Kopper, B. A., Baker, M. T., & Haraburda, C. M. (2002). Development of the reasons for living inventory for young adults. *Journal of Clinical Psychology, 58*(4), 339–357. doi:10.1002/jclp.1147.

Gutierrez, P. M., Osman, A., Kopper, B. A., & Barrios, F. X. (2000). Why young people do not kill themselves: The Reasons for Living Inventory for Adolescents. *Journal of Clinical Child Psychology, 29*(2), 177–187. doi:10.1207/S15374424jccp2902_4.

Guy, L. S., Packer, K., & Warnken, W. (2012). Assessing risk of violence using structured professional judgment guidelines. *Journal of Forensic Psychology Practice, 12*(3), 270–283. doi:10.1080/15228932.2012.674471.

Haney, E. M., O'Neil, M. E., Carson, S., Low, A., Peterson, K., Denneson, L. M., Kansagara, D. (2012). Suicide risk factors and risk assessment tools: a systematic review. *VA-ESP Project #05-225.* https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0042005/pdf/PubMedHealth_PMH0042005.pdf. Accessed 24 May 2017.

Hart, S. D. (1998a). Psychopathy and risk for violence. In D. Cooke, A. E. Forth & R. D. Hare (Eds.), *Psychopathy: theory, research, and implications for society* (pp. 355–375). Dordrecht: Kluwer.

Hart, S. D. (1998b). The role of psychopathy in assessing risk for violence: conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121–137. doi:10.1111/j.2044-8333.1998.tb00354.x.

Hawton, K., Rodham, K., Evans, E., & Weatherall, R. (2002). Deliberate self-harm in adolescents: self report survey in schools in England. *BMJ (Clinical Research Ed.), 325*, 1207–1211.

Heilbrun, K. (2012). *Developments in risk assessment: Violence risk and sexual violence risk*. Report for the American Psychology-Law Society, San Juan, PR. http://www.drexel.edu/~/media/Files/psychology/labs/heilbrun/presentations/RAWS%20APLS%20San%20Juan%203%2012%20WORD.ashx. Accessed 15 May 2017.

Heudo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *CHIP Documents*. Paper 19. http://digitalcommons.uconn.edu/cgi/viewcontent.cgi?article=1019&context=chip_docs. Accessed 22 May 2017.

Hilterman, E. L., Nicholls, T. L., & van Nieuwenhuizen, C. (2014). Predictive validity of risk assessments in juvenile offenders: Comparing the SAVRY, PCL:YV, and YLS/CMI with unstructured clinical assessments. *Assessment, 21*(3), 324–339. doi:10.1177/1073191113498113.

Hoge, R. D., & Andrews, D. A. (2002). *The youth level of service/case management inventory manual and scoring key*. North Tonawanda, NY: Multi Health Systems.

Jessor, R. (2014). Problem behavior theory: a half-century of research on adolescent behavior and development. In R. M. Lerner, A. C. Petersen, R. K. Silbereisen & J. Brooks-Gunn (Eds.), *The developmental science of adolescence: History through autobiography*. New York and London: Psychology Press.

Jessor, R., Donovan, J. E., & Costa, F. M. (1991). *Beyond adolescence: Problem behavior and young adult development*. New York: Cambridge University Press.

Jessor, R., & Turbin, M. S. (2014). Parsing protection and risk for problem behavior versus pro-social behavior among U.S. and Chinese adolescents. *Journal of Youth and Adolescence, 43*(7), 1037–1051. doi:10.1007/s10964-014-0130-y.

Koo, H. J., & Kwon, J.-H. (2014). Risk and protective factors of internet addiction: A meta-analysis of empirical studies in Korea. *Yonsei Medical Journal, 55*, 1691–1711. doi:10.3349/ymj.2014.55.6.1691.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. doi:10.3389/fpsyg.2013.00863.

Langbehn, D. R., Cadoret, R. J., Yates, W. R., Troughton, E. P., & Stewart, M. A. (1998). Distinct contributions of conduct and oppositional defiant symptoms to adult antisocial behavior. *Archives of General Psychiatry, 55*, 821–829. doi:10.1001/archpsyc.55.9.821.

Laub, J. H., & Sampson, R. J. (1991). The Sutherland-Glueck debate: On the sociology of criminal knowledge. *The American Journal of Sociology, 96*, 1402–1440. doi:10.1086/229691.

Lodewijks, H. P., de Ruiter, C., & Doreleijers, T. A. (2008a). Gender differences in violent outcome and risk assessment in adolescent offenders after residential treatment. *The International Journal of Forensic Mental Health, 7*(2), 133–146. doi:10.1080/14999013.2008.9914410.

Lodewijks, H. P., Doreleijers, T. A., & de Ruiter, C. (2008b). SAVRY risk assessment in violent Dutch adolescents: Relation to sentencing and recidivism. *Criminal Justice and Behavior, 35*(6), 696–709. doi:10.1177/0093854808316146.

Lodewijks, H. P., Doreleijers, T. A., de Ruiter, C., & Borum, R. (2008c). Predictive validity of the structured assessment of violence risk in youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry, 31*(3), 263–271. doi:10.1016/j.ijlp.2008.04.009.

Lodewijks, H. P. B., de Ruiter, C., & Doreleijers, T. A. H. (2010). The impact of protective factors in desistance from violent reoffending: a study in three samples of adolescent offenders. *Journal of Interpersonal Violence, 25*(3), 568–587. doi:10.1177/0886260509334403.

Loeber, R., & Farrington, D. P. (1998). Never too early, never too late: Risk factors and successful interventions for serious and violent juvenile offenders. *Studies on Crime and Crime Prevention, 7*, 7–30.

Madge, N., Hewitt, A., Hawton, K., de Wilde, E. J., Corcoran, P., Fekete, S., & Ystgaard, M. (2008). Deliberate self-harm within an international community sample of young people: comparative findings from the Child and Adolescent Self-harm in Europe (CASE) Study. *Journal of Child Psychology and Psychiatry, 49*, 667–677. doi:10.1111/j.1469-7610.2008.01879.x.

Manor, I., Gutnik, I., Ben-Dor, D. H., Apter, A., Sever, J., Tyano, S., & Zalsman, G. (2010). Possible association between attention deficit hyperactivity disorder and attempted suicide in adolescents—a pilot study. *European Psychiatry, 25*, 146–150. doi:10.1016/j.eurpsy.2009.06.001.

McEachran, A. (2001). *The predictive validity of the PCL: YV and the SAVRY in a population of adolescent offenders*. Burnaby, BC: Simon Fraser University. http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp04/MQ61591.pdf. Accessed 23 May 2017

McGowan, M. R., Horn, R. A., & Mellott, R. N. (2011). The predictive validity of the tructured Assessment of Violence Risk in Youth in secondary educational settings. *Psychological Assessment, 23*(2), 478–486. doi:10.1037/a0022304.

Meyers, J. R., & Schmidt, F. (2008). Predictive validity of the structured assessment for violence risk in youth (SAVRY) with juvenile offenders. *Criminal Justice and Behavior, 35*(3), 344–355. doi:10.1177/0093854807311972.

Miccio-Fonseca, L. C. (2009). MEGA ♪: A new paradigm in protocol assessing sexually abusive children and adolescents. *Journal of Child and Adolescent Trauma, 2*, 124–141. doi:10.1080/19361520902922434.

Miccio-Fonseca, L. C. (2013). MEGA ♪: A new paradigm in risk assessment tools for sexually abusive youth. *Journal of Family Violence, 28*(6), 623–634. doi:10.1007/s10896-013-9527-8.

Miccio-Fonseca, L. C. (2016). MEGA ♪ cross-validation findings on sexually abusive females: Implications for risk assessment and clinical practice. *Journal of Family Violence*. doi:10.1007/s10896-016-9845-8.

Miccio-Fonseca, L. C., & Rasmussen, L. (2013). Applicability of MEGA ♪ to sexually abusive youth with low intellectual functioning. *Journal of Mental Health Research in Intellectual Disabilities, 6*(1), 42–59. doi:10.1080/19315864.2011.650788.

Miccio-Fonseca, L. C., & Rasmussen, L. (2015). MEGA♪—empirical support for nomenclature on the anomalies: Sexualy violent and predatory youth. *International Journal of Offender Therapy and Comparative Criminology, 59*(11), 1222–1238. doi:10.1177/0306624X14533265.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. for the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ, 339*, b2535. doi:10.1136/bmj.b2535.

National Center for Juvenile Justice (2012). *Models for change update 2012: Headlines*. Pittsburgh, PA: National Center for Juvenile Justice. http://www.ncjj.org/Publication/Models-for-Change-Update-2012-Headlines.aspx. Accessed 25 May 2017.

O' Shea, L. E., Mitchell, A. E., Picchioni, M. M., & Dickens, G. L. (2013). Moderators of the predictive efficacy of the historical, clinical and risk management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. *Aggression and Violent Behavior, 18*, 255–270. doi:10.1016/j.avb.2012.11.016.

O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the short-term assessment of risk and treatability (START) for aggression and self-harm in a secure mental health service: gender differences. *International Journal of Forensic Mental Health, 14*, 132–146. doi:10.1080/14999013.2015.1033112.

O'Shea, L. E., & Dickens, G. L. (2016). Performance of protective factors assessment in risk prediction for adults: systematic review and meta-analysis. *Clinical Psychology Science and Practice, 23*, 126–138. doi:10.1111/cpsp.12146.

Olver, M. E., Olver Keira, C., Stockdale, J., & Wormith, S. (2009). Risk assessment with young offenders: a meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36*, 329–353. doi:10.1177/0093854809331457.

Osman, A., Kopper, B. A., Barrios, F. X., Osman, J. R., Besett, T., & Linehan, M. M. (1996). The brief reasons for living inventory for adolescents (BRFL-A). *Journal of Abnormal Child Psychology, 24*(4), 433–443. doi:10.1007/bf01441566.

Owens, T. (2011). *Utility of the SAVRY in Predicting Recidivism Among Juvenile Sex Offenders*. http://docs.rwu.edu/cgi/viewcontent.cgi?article=1017&context=psych_thesis. Accessed 23 May 2017. **(National Criminal Justice Reference Service (NCJRS) Abstracts Database database)**.

Patton, G. C., Coffey, C., Sawyer, S. M., Viner, R. M., Haller, D. M., Bose, K., & Mathers, C. D. (2009). Global patterns of mortality in young people: a systematic analysis of population health data. *Lancet, 374*, 881–892. doi:10.1016/S0140-6736(09)60741-8.

Penney, S. R., Lee, Z., & Moretti, M. M. (2010). Gender differences in risk factors for violence: An examination of the predictive validity of the structured assessment of violence risk in youth. *Aggressive Behavior, 36*(6), 390–404. doi:10.1002/ab.20352.

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., Mann, J. (2011). The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *The American Journal of Psychiatry, 168*(12), 1266–1277. doi:10.1176/appi.ajp.2011.10111704.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rennie, C. E., & Dolan, M. C. (2010). The significance of protective factors in the assessment of risk. *Criminal Behavior and Mental Health, 20*(1), 8–22. doi:10.1002/cbm.750.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior, 29*, 615–620. doi:10.1007/s10979-005-6832-7.

Richard, K. M. (2013). Predicting persistence and desistence of recidivism in youth offenders: The role of risk and protective factors in criminal offending. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 73*(8-B(E)), No Pagination Specified.

Rolf, J. E. (1985). Evolving adaptive theories and methods for prevention research with children. *Journal of Consulting and Clinical Psychology, 53*, 631–646 **(Special section: special series: primary prevention of childhood disorders)**.

Rosenberg, M. S. (1987). New directions for research on the psychological maltreatment of children. *American Psychologist, 42*, 166–171.

Schmidt, F., Campbell, M. A., & Houlding, C. (2011). Comparative analyses of the YLS/CMI, SAVRY, and PCL:YV in adolescent offenders: a 10-year follow-up Into adulthood. *Youth Violence and Juvenile Justice, 9*(1), 23–42. doi:10.1177/1541204010371793.

Shader, M. (2001). *Risk factors for delinquency: An overview*. US Department of Justice Office of Juvenile Justice and Delinquency Prevention. https://www.ncjrs.gov/pdffiles1/ojjdp/frd030127.pdf?q=risk-and-protective-factors-of-child-delinquency. Accessed 07 May 2017.

Shepherd, S. M., Luebbers, S., Ferguson, M., Ogloff, J. R., & Dolan, M. (2014a). The utility of the SAVRY across ethnicity in Australian young offenders. *Psychology, Public Policy, and Law, 20*(1), 31–45. doi:10.1037/a0033972.

Shepherd, S. M., Luebbers, S., Ogloff, J. R. P., Fullam, R., & Dolan, M. (2014b). The predictive validity of risk assessment approaches for young Australian offenders. *Psychiatry, Psychology and Law, 21*, 801–817. doi:10.1080/13218719.2014.904262.

Silver, E., & Miller, L. L. (2002). A cautionary note on the use of actuarial risk assessment tools for social control. *Crime and Delinquency, 48*, 138–161.

Singh, J. P., Serper, M., Reinharth, J., & Fazel, S. (2011). Structured assessment of violence risk in schizophrenia and other psychiatric disorders: a systematic review of the validity, reliability, and item content of 10 available instruments. *Schizophrenia Bulletin, 37*, 899–912. doi:10.1093/schbul/sbr093.

Teplin, L. A., Abram, K. M., McClelland, G. M., Mericle, A. A., Dulcan, M. K., & Washburn, J. J. (2006). *Psychiatric disorders of youth in detention*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Turner, S., & Fain, T. (2006). Validation of the risk and resiliency assessment tool for juveniles in the Los Angeles county probation system. *Federal Probation, 70*(2), 49–57.

United States Department of Justice (2017) *Crime in the U.S. 2015 [Table 5]*. Federal Bureau of Investigation Criminal Justice Information Services Division. https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015. Accessed 07 May 2017.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*. doi:10.18637/jss.v036.i03.

Viljoen, J. L., Beneteau, J. L., Gulbransen, E., Brodersen, E., Desmarais, S. L., Nicholls, T. L., & Cruise, K. R. (2012). Assessment of multiple risk outcomes, strengths, and change with the START:AV: A short-term prospective study with adolescent offenders. *The International Journal of Forensic Mental Health, 11*(3), 165–180. doi:10.1080/14999013.2012.737407.

Viljoen, J. L., Cruise, K. R., Nicholls, T. L., Desmarais, S. L., & Webster, C. D. (2012). Taking stock and taking steps: The case for an adolescent version of the short-term assessment of risk and

treatability. *International Journal of Forensic Mental Health, 11*(3), 135–149. doi:10.1080/14999013.2012.737406.

Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chavez, V., Ullman, D., & Lawrence, L. (2008). Assessing risk for violence in adolescents who have sexually offended. *Criminal Justice and Behavior, 35*(1), 5–23. doi:10.1177/0093854807307521.

Vincent, G. M., Perrault, R. T., Guy, L. S., & Gershenson, B. G. (2012). Developmental issues in risk assessment: Implications for juvenile justice. *Victims and Offenders, 7*(4), 364–384. doi:10.1080/15564886.2012.713900.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the short-term assessment of risk and treatability (START) (Version 1.1). Coquitlam, BC*. Canada: British Columbia Mental Health and Addiction Services.

Webster CD, Douglas KS, Eaves D, Hart SD (1997) *HCR-20: Assessing Risk for Violence (Version 2)*. Burnaby, BC: Mental Health, Law and Policy Institute, SimonFraser University.

Welsh, J. L., Schmidt, F., McKinnon, L., Chattha, H., & Meyers, J. R. (2008). A comparative study of adolescent risk assessment instruments: Predictive and incremental validity. *Assessment, 15*(1), 104–115. doi:10.1177/1073191107307966.

Westefeld, J. S., Cardin, D., & Deaton, W. L. (1992). Development of the college student reasons for living inventory. *Suicide and Life-Threatening Behavior, 22*(4), 442–452.

Wheaton, B. (1986). Models for the stress-buffering functions of coping resources. *Journal of Health and Social Behavior, 26*, 352–365.

Worling, J. R. & Curwen, C. (2001). *The "ERASOR": Estimate of Risk of Adolescent Sexual Offense Recidivism Version 2.0*. Toronto: SAFE-T Program, Thistletown Regional Centre. http://djj.ky.gov/800%20Policy%20Manual/ERASOR%202.0.pdf. Accessed 07 May 2017.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics, 42*, 565–569. doi:10.1375/twin.13.6.517.