**How effective are risk assessments/measures for predicting future aggressive behaviour in adults with intellectual disabilities (ID): A systematic review and meta-analysis.**

Corresponding author: Dr Rachael Lofthouse, The Harbour, Blackpool, Preston New Road, Blackpool FY4 4FE.
Rachael.lofthouse@yahoo.co.uk
Tel: 07793 144 959

Co – authors:
Dr Laura Golding, University of Liverpool
Dr Vasiliki Totsika and Professor Richard Hastings, CEDAR and Centre for Education Studies (CES), University of Warwick; Centre for Developmental Psychiatry and Psychology, Department of Psychiatry, Monash University.and Professor William Lindsay[1], The Danshell Group.

[1] Since this manuscript was originally accepted, sadly our colleague Bill Lindsay has passed away.

Aggression directed toward others and the environment is one of the most difficult to manage behaviours in services for adults with intellectual disabilities (ID) (Ali, Hall, Blickwedel, & Hassiotis, 2015). Prevalence rates of aggression for this population vary from 9.2% (Borthwick-Duffy, 1994) to 51% (Crocker et al., 2006) due to the methodological variations across studies. Aggressive behaviour threatens the safety and well being of the adult with ID as well as carers and others around the individual. The emotional and psychological well being of staff is negatively impacted (Hastings, 2002; Hastings & Brown, 2002) causing stress and burnout (Hensel, Lunsky, & Dewa, 2012; Mills & Rose, 2011). For the individual with ID, exhibiting aggression leads to an increased likelihood of being excluded from services, a negative impact on self-esteem, and restriction of the ability to maintain social networks (Cooper et al., 2009).

Risk assessment is "the process of evaluating individuals to characterise the risk that they will commit aggression in the future, and to develop interventions to manage or reduce that risk" (Boer et al, 1997). Thus, accurate assessment of risk is considered to be essential for successfully reducing risk (Campbell, French, & Gendreau, 2009). Structured clinical assessments are considered to be the optimal method for systematically assessing risk of aggression (Monahan et al., 2001). This view is supported by guidelines in the UK (Department of Health, 2009; Royal College of Psychiatrists, 2007; National Institute for Health and Clinical Excellence guidelines; NICE, 2015).

The Risk, Need, Responsivity (RNR) model outlined by Andrews and Bonta (2006) has influenced the development of many risk assessment measures in ID and non-ID populations. According to the model, the *risk* principle stipulates that the behaviour of interest, such as aggression, can be reliably predicted and that treatment

resources should focus on higher risk individuals. This process typically involves actuarial risk assessments; largely focused on static risk factors. Although actuarial measures have demonstrable efficacy in predicting violent recidivism (Hanson & Morton-Bourgon, 2007; 2009), the clinical contribution to the day-to-day management of risk is limited. Aggregate data might not translate to individual cases (Doyle & Dolan, 2007) failing to comprehensively capture the individual's unique circumstances (Hart & Cooke, 2013).

The *need* principle relates to 'criminogenic need' or 'dynamic' risk factors that are psychological or behavioural features of the individual that can be influenced and changed by psychological, social or physiological variables (Wong & Gordon, 2006). Dynamic factors are also amenable to deliberate intervention or change (risk decrease). Structured Professional Judgement (SPJ) measures typically include historical items that are fixed (static) alongside dynamic risk factors (e.g., pro-criminal attitudes). The SPJ measures are proposed as a useful addition or alternative to the actuarial approach (Hart & Logan, 2011) for the day-to-day management of risk. The *responsivity* principle describes how treatment should be tailored to the individual's motivation, ability and learning style to maximise success.

**Risk assessment approaches**

Beech and Ward (2004) offer an alternative method of conceptualising risk factors to the traditional static/dynamic split. The authors propose that historical factors (static) act as a marker for psychological meaningful risk factors (dynamic). For example, a history of violent behaviour may be indicative of a current anti social attitude. In their analysis of the performance of various static and dynamic risk measures, Lofthouse and colleagues (2014a) found that static and dynamic

approaches tapped into the same underlying risk. Specifically, dynamic measures can act as a 'proxy' for static measures. Lofthouse et al. concluded that dynamic risk measures may be more meaningful than static measures for assessing risk in an ID population. Other authors (on the basis of conceptual or clinical reasoning) call for a convergent approach (Boer, Tough, & Haaven, 2004; Pouls & Jeandarme, 2015). From this perspective, static risk measures are used to establish a 'risk baseline' and inform treatment intensity and supervision levels. Dynamic measures are employed to assess, identify and monitor change in targets for treatment (Pouls & Jeandarme, 2015).

Emerging research in the ID field highlights a link between dynamic risk factors (e.g., lack of structured routine activity and the quality of close relationships) and an increased risk of offending (Wheeler, Clare, & Holland, 2014). In line with the RNR model (Andrews & Bonta, 2006) researchers recognise that due to their unique needs, adults with ID are likely to be more interdependent within services. Therefore, factors relating to the environment (e.g., staff knowledge of the individual) are equally important as those relating to the individual (e.g., historical, dispositional) for a comprehensive and ecologically valid assessment of risk (Boer et al., 2004; Lofthouse et al., 2013). The ID field has seen a steady increase in risk assessment measures developed for this population following this approach (Boer et al., 2004; Lindsay et al., 2008; Lofthouse et al., 2014b; Quinsey, Book, & Skilling, 2004).

The uncertainty regarding the reliability and validity of using non ID-specific methods of assessing risk with adults with ID, the unique characteristics of people with ID, and the paucity of research and empirically supported risk assessments in the ID field, leaves many researchers, policy makers and professionals uncertain which assessment to use in research and clinical practice. This is concerning given the

evidence demonstrating that aggression poses a major challenge for service users with ID, their carers and service providers (Cooper et al., 2009). The situation highlights the significant need for a systematic and comprehensive synthesis of the evidence on the predictive accuracy of existing methods.

**Previous reviews**

We know from meta-analytic studies in the non-ID field (Campbell, French & Gendreau, 2009; Fazel, Singh, Doll & Grann, 2012; Singh, Grann & Fazel, 2011; Yang, Wong & Coid, 2010;) that actuarial and SPJ approaches have similar accuracy, generally in the moderate to high range, for predicting violence across different samples (Mills, 2017). Campbell and colleagues (2009) in their meta-analysis of 88 studies, found that although the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice & Cormier, 2006) had the highest predictive accuracy, the effect sizes for both types of measures (static/dynamic) did not differ significantly and that overall, dynamic measures performed better than static measures. Furthermore, Singh and colleagues (2011) found that measures designed specifically to assess violence (i.e., HCR-20, VRAG, SARA [Spousal Assault Risk Assessment; Kropp, Hart, Webster & Eaves, 1999] and SAVRY [Structured Assessment for Violence Risk in Youth; Borum, Bartel & Forth, 2006]) had greater accuracy than those used to predict general or sexual reoffending.

To date, there is no meta-analytic study conducted with ID samples. Therefore, it is unclear which of the two approaches is better for risk prediction. Three recent narrative reviews have been conducted (Camilleri & Quinsey, 2011; Hockenhull, n.d.; Pouls & Jeandarme, 2015) with the aim of providing guidance for professionals in the selection and interpretation of risk assessments for individuals with ID. However, the focus on retrospective designs, the inclusion of general

violence and sexual violence outcomes and a lack of robust quantitative synthesis hampers the confidence that can be placed in the recommendations and highlights the need for further research evidence.

The aim of the present study was to provide a systematic review and meta-analysis of existing evidence on the predictive validity of available methods of predicting risk of aggression among individuals with ID. The current study aimed to address limitations in the research evidence so far by: (a) establishing more stringent criteria for ID (e.g., IQ <70) to accurately define the sample, because in ID forensic services there is a clinical need to understand what works most effectively for adults with ID as a distinct group, (b) focusing on aggression only and not sexual offences given the evidence for different risk factors relating to different types of offending (Craig et al., 2013), (c) including only prospective studies considered to be higher quality and thus generating more accurate results, and (d) including a meta-analysis to synthesize findings in a summary statistic that is useful to guide clinical decision-making. Meta-analysis is considered the most robust method of synthesizing from quantitative research studies.

## Method

### Review protocol

To ensure consistency, the current review followed the guidance set out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (Moher, Liberati, Tetzlaff, & Altman, 2009). In line with this guidance, an unpublished systematic review protocol was developed to comprehensively and objectively search the literature (available on request from the first author).

**Search strategy**

A systematic literature search was conducted using the following databases: PsycINFO, EMBASE, MEDLINE, Web of Science, and Google Scholar. Given the limited research in this area, the search was not restricted by date. Only articles published in English were included. Studies were identified by combining search terms specifying a sample with intellectual disabilities (i.e. intellectual disab*, learning disab*, developmental disab*, mental retard*), terms specifying risk assessment (risk AND assessment, risk AND management, risk AND prediction, risk AND measure, risk AND tool), terms were used to restrict the search to studies with aggression as the outcome variable (Violen*, aggressi*, challenging behavio*; NOT sexual AND violen*, aggressi*) and prospective studies only (NOT retrospective). Additional empirical studies were identified through review of the reference list of articles collected in the search described above. An email request was also sent to 43 international researchers known to conduct research in the field to obtain any unpublished or in press studies.

**Study selection**

From this initial search, eligibility for inclusion in the meta-analysis was determined by the following inclusion/exclusion criteria:

**Population**. Adults (aged 18 years and above) identified as having an intellectual disability or equivalent diagnosis (e.g., learning disability in the UK, mental retardation or developmental delay) using any one of the following criteria: IQ < 70, as assessed with standardised measures; impairments in adaptive behaviour assessed with adaptive behaviour scales; or administratively defined as currently receiving ID services.

***Risk Assessment***. Risk assessments were defined as structured and standardized measures containing one or more factors considered to be predictive of verbal or physical aggression. Such measures could include: Structured Professional Judgment (SPJ), Actuarial risk assessment, Static risk assessment, Dynamic risk assessment, measures combining one or more of the above approaches, or measures adapted for ID populations such as the HCR-20 ID supplement, measures developed for other purposes such as personality assessments (e.g., PCL-R).

***Outcome***. The likelihood of verbal, physical aggression or both. There is no universally accepted definition of aggression (Yang, Wong, & Coid, 2010). For the purpose of this review, physical aggression is defined as an act of physical violence, aggression, or force with hostility and intention to hurt or damage someone or something physically or psychologically (Yang et al., 2010). Verbal aggression is defined as having content that is threatening, hostile or derogatory; aimed at a specific individual or individuals and would be perceived as causing offence because of its content and/or severity/intensity. Aggression charges or convictions as well as noncriminal aggression toward persons or environment were included. The decision was taken to exclude self-injurious behaviour and sexual aggression from this review because of the potentially different and complex aetiology of these behaviours. Sexual and non-sexual aggression is commonly thought to have different causes and antecedents (Lim & Howard, 1998). Outcome measures covering a variety of domains were included if the aggression outcome (e.g., sub-scale) was reported separately. Measures of attitudes/beliefs relating to sexual aggression where no physical/verbal aggression was measured were also excluded from the current study.

***Study type***. Studies were included in this review that were prospective in design and included a minimum follow up period of one day. Any relevant design was accepted:

cohort studies, randomized control trials, case-control studies, and experimental case studies. Catch up longitudinal design studies were included where follow up data could potentially have been collected concurrently to the administration of the risk assessment. This is a common approach in risk assessment research.

***Setting***. No restrictions were imposed on setting (e.g., community, mental health, forensic).

The study selection process is summarized in Figure 1 using a PRISMA flowchart. The electronic and manual searches resulted in 595 potential hits. All titles and abstracts were reviewed by the first author (RL) using the inclusion/exclusion criteria outlined above. A second person (RF) was available to discuss more ambiguous studies. All title and abstracts of identified citations were reviewed by a second reviewer (PhD student in Psychology). At this stage, no further studies were identified for inclusion. The 47 studies that remained after this initial screen were coded by the first author and cross-checked with the second reviewer and to make final inclusion decision.

<Insert figure 1>

**Quality assessment**

Methodological quality of the included studies was assessed using the Critical Appraisal Skills Program Cohort Study Checklist (CASP, 2013). This method comprises a checklist of 9 items. Items were rated on a three-point scale: 2 (criterion present), 1 (partially present), and 0 (absence of the criteria or insufficient information). Two items (confounding variables) were omitted from the assessment because Receiver Operating Characteristic (ROC) Area Under Curve (AUC) analysis does not require a multivariate analysis. Item scores were summed to produce an

overall quality score; higher scores (maximum possible score = 18) were indicative of better quality (Table 1). Studies were generally of high quality, within the range 12 – 17 (mean score = 13.5). Some risk of bias was apparent for four studies due to limited information regarding the method used to recruit participants within study sites and unclear criteria for definition of ID. Furthermore, four studies failed to adequately operationalize the term 'violence/aggression'.

**Data extraction**

Information for each study was extracted on sample size, participant gender and age, level of ID, and outcome data. Two variables were coded for subsequent subgroup analysis: study design (prospective vs. catch-up longitudinal) and type of measure (static, SPJ, dynamic). Outcome statistics obtained from studies were AUC, standard error (SE), confidence intervals (CI), and correlations.

**Statistical analysis**

A meta-analysis was undertaken of reported AUCs to produce a single summary AUC estimate, weighted by the inverse of study variance. Rice and Harris (2005) offer the following Cohen's *d* effect size equivalent for AUC: small (.556), medium (.639) and large (.714). The meta-analysis used AUCs as reported in the primary studies, or if studies reported correlation coefficients, these were converted to AUCs. This conversion followed available guidance from Zhou and colleagues (2002). Where missing in studies, standard errors were obtained from confidence intervals and *p* values (Higgins, Thompson, Deeks, & Altma, 2003). The meta-analysis was conducted using MedCalc® Software (Schoonjans, Zalata, Depuydt, & Comhaire, 1995).

## Tests of homogeneity and publication bias

To determine whether all studies were drawn from a population of studies with a common main effect size, we performed a test of homogeneity using the Q-statistic and $I^2$, utilizing these options in MedCalc® software. These tests were conducted on the whole group of 14 studies. In addition, we assessed potential for publication bias by a funnel plot of the standard error and effect size for each study (Egger, Smith, Schneider, & Minder, 1997).

# Results

## Description of studies

Table 1 outlines the study characteristics of the 14 included studies. A total of 1,390 participants were included across all studies. The average number of participants per study was 99.29 ranging between 23 and 218. The majority of participants were male, with only two studies including female participants. The mean age of participants across studies was 36.39, mean ages across studies ranged from 29.77 to 41.9 years.

For those studies that reported IQ data (n=9), the average IQ was 65.16. Three studies reported classification of ID using the ICD 10 Mental Retardation (F70—F79; Gray, Fitzgerald, Taylor, MacCulloch, & Snowden, 2007; Gray, Taylor, & Snowden, 2011; O'Shea, Picchioni, Mason, Sugarman, & Dickens, 2015). One study reported that participants had mental retardation (Quinsey et al., 2004), one study administratively defined participants as having an ID by virtue of receiving ID services (Lofthouse et al., 2014b). In one study, level of ID was unspecified (Fitzgerald et al., 2011) and one study reported presence of 'learning disability' (within UK services) ranging from borderline to moderate (Innet, Wright, Roberts, &

Sheeran, 2014).

Eight studies were prospective studies and six were catch-up longitudinal prospective. Where stated, the follow up time in prospective studies ranged from three months to five years. The majority of studies (n =11) were conducted in forensic high or medium secure settings. The remaining studies (some included multiple settings) were conducted in low secure, rehabilitation, acute or inpatient mental health settings; prison or community settings following discharge from medium secure settings (n = 9).

The majority of studies included actuarial measures to assess the risk of aggression. Four studies: Quinsey et al. (2004); Gray et al. (2007); Lindsay et al. (2008) and Fitzgerald et al. (2013) included the Violence Risk Appraisal Guide [VRAG]; Quinsey, Harris, Rice, & Cormier, 1998, 2006), Fitzgerald et al. (2011) used the Offender Group Reconviction Scale [OGRS]; two studies: Gray et al. (2007) and Pouls and Jeandarme (2014) used the Psychopathy Checklist – Screening Version [PCL-SV]; three studies: Morrissey et al. (2005; 2007) Pouls and Jeandarme (2014) used the Psychopathy Checklist-Revised [PCL-R] Hare (2003). In relation to SPJ, four studies: Gray et al. (2007); Lindsay et al. (2008); Fitzgerald et al. (2013); and O'Shea et al. (2015) assessed risk using the Historical Clinical Risk Management – 20 [HCR20] Douglas, Hart, Webster, and Belfrage (2013). The remaining studies assessed risk using dynamic assessments, three studies: Morrissey et al. (2005; 2007) and Lindsay et al. (2008) focused on the Emotional Problem Scale-Behaviour Rating Scale [EPS-BRS] Prout and Strohmer (1991). The following assessments were all assessed in one study each: Quinsey et al. (2004) used the Problem Identification Checklist [PIC] Quinsey, Coleman, Jones, and Altrows (1997); Quinsey et al. (2004) assessed risk using the Proximal Risk Factor Scale [PRFS] Quinsey, et al. (1997);

Steptoe, Lindsay, Murphy, and Young (2008) used Dynamic Risk Assessment and Management System [DRAMS] Lindsay et al. (2004); Drieschner, Marrozo, and Regenboog (2013) included Dynamic Risk Outcome Scale [DROS] Drieschner and Hesper (2008); Innet et al. (2014) included Short-Term Assessment of Risk and Treatability [START] Webster, Martin, Brink, Nicholls, and Middleton (2004); Lindsay et al. (2008) used the Short Dynamic Risk Scale [SDRS] Quinsey (2004) and Lofthouse et al. (2014b) included the Current Risk of Violence [CuRV] Lofthouse et al. (2014b).

## Meta-analysis

### Homogeneity and publication bias

The Q statistic was statistically significant $Q(14) = 46.53$, $p < .01$, for scores across the studies. The results suggested that there was significant heterogeneity between the studies. We also calculated the between study variance ($I^2 = 72.06$) and these data supported the homogeneity conclusion in that relatively large proportions of variance were explained by between study variance. To address this, we used the random effects approach to the calculation of the summary effect size. We attempted to explore sources of heterogeneity through planned subgroup analyses.

We found no statistical or visual evidence of publication bias. Figure 2 shows a funnel plot of standard error against AUC effect size of studies. However, given the limitation of this technique when a small number of studies are included, we cannot rule out the possibility of publication bias.

<Insert figure 2>

**Effect size measures**

Meta-analysis was conducted using MedCalc® statistical software (Schoonjans et al., 1995). Effect sizes were computed for each individual study. Where more than one relevant AUC was reported in one study, the mean was calculated. The test of homogeneity suggested heterogeneity and for this reason we estimated the summary weighted effect size using a random-effects approach. The summary weighted effect size from all studies (n=14) suggested a significant medium to large effect size within the confidence intervals (AUC= .724, 95% CI: 0.681, 0.768) (see Forrest Plot in Figure 3 for effect size and confidence intervals for the 14 included studies). The large standard CIs found in the Steptoe et al. (2008) and Morrissey et al. (2005) study may either be because of a non-specific effect or because of measurement variability that might have been caused by the formula conversions when transforming a correlation coefficient to an AUC.

<Insert Figure 3>

**Subgroup analyses**

**Type of risk assessment**

A secondary aim was to explore whether overall effectiveness is likely to be moderated by the type of risk assessment used (i.e., actuarial vs. SJP vs. dynamic), or by study design (catch-up longitudinal vs. prospective). The effect size and 95% CIs for type of risk assessment are shown in Table 2.

<Insert Table 2>

The results suggest that all three types of measures predict aggression at a level significantly better than chance (AUC= 0.5) and all three measures have a large effect size (Rice & Harris, 2005). The overlapping confidence intervals for the three methods do not suggest that there are significant differences between the three types

of risk assessments. There is an indication that the measures provide adequate prediction according to their effect sizes (see Figures 4 – 6 for Forrest Plots for mean effect size and CIs for studies including actuarial, SPJ and dynamic measures respectively). The same study may appear in different forest plots if multiple measures are used within the study. The AUC reported is the relevant measure in that study (i.e., Actuarial, SPJ, Dynamic).

<Insert Figures 4-6>

**Study Design**

The overall effect size and CIs for design of study is shown in Table 3.

<Insert Table 3>

This result suggests that studies that use a catch-up longitudinal design have a large effect size whilst prospective designs have a medium effect size. However, the AUCs are very similar and the confidence intervals of the two types of studies do not suggest that there are significant differences between the two study designs. See Forrest Plot in Figure 7 for effect size and confidence intervals for the studies using a prospective design and Figure 8 for the studies utilising a catch-up longitudinal design.

<Insert Figures 7 & 8>

## Discussion

The current study aimed to synthesise available evidence relating to the effectiveness of risk assessment measures for predicting risk of aggression in individuals with ID. The summary weighted effect size was large and significant,

indicating that available risk assessments measures can predict future aggression significantly better than chance. Findings from the current study are in line with previous narrative reviews (Camilleri & Quinsey, 2011; Hockenhull, n.d.; Pouls & Jeandarme, 2015) that found evidence for the predictive validity of several risk assessment measures for males with ID and a history of offending behaviour. The present meta-analysis expands and improves previous studies by conducting a meta-analysis to synthesise findings in a summary statistic that is useful to guide clinical decision making. Unlike the previous systematic and narrative reviews in this area, the present meta-analysis focused exclusively on risk of aggression and included only prospective studies in an attempt to improve our understanding of specific methods of assessment and minimise biases.

To explore potential moderators of effectiveness, the effect of type of risk assessment measure on predictive accuracy was examined. On the basis of evidence from the current study, and in line with some previous studies in the general offender literature (Singh & Fazel, 2010; Wong et al., 2010) and ID literature (Fitzgerlad et al., 2013; Gray et al., 2007; Lindsay et al., 2008) there does not seem to be a difference in the prediction between the three types of measures. In the present study, based on the magnitude of the effect size, dynamic measures significantly predicted risk, but they did so slightly less well than actuarial and SPJ methods. This finding supports the preponderance of actuarial or SPJ approaches in assessing risk in practice. The caveat, however, is that at this stage, the evidence regarding dynamic measures was not directly comparable with regard to methodological aspects for the other two types of measures. These findings might be due to variation and methodological quality of the scales included within the dynamic measure group. The actuarial and SPJ groups included studies that used only the same scale (i.e. VRAG and HCR 20, for actuarial

and SPJ, respectively). These risk measures were developed specifically to measure risk of violence/aggression (albeit among mainstream offenders). However, the dynamic subgroup included a wider variety of measures (CuRV, EPS-BRS, DRAMS, SDRS, DROS, START). Some of the measures in the dynamic subgroup (e.g., EPS-BRS) had not been originally developed with the intention of assessing risk in any population. Other measures have not been subjected to extensive research evaluation and therefore do not have established psychometric properties. For example, the study by Lofthouse and colleagues (2014b) included in the present meta-analysis, was the only piece of research assessing the efficacy of the CuRV dynamic risk measure. The heterogeneity and diversity of dynamic measures currently being used within ID settings is likely due to the unavailability of measures given that the research in this area is at the early stage of development. Researchers and clinicians hampered by the lack of measures commonly produce their own (e.g., CuRV and DRAMS) informed by their own clinical experience and research evidence. Whereas other studies have included measures frequently used within their clinical practice (EPS-BRS).

The present study also explored the potential moderating effect of study design. Findings suggested there was no difference in the prediction of risk between catch-up and other prospective study designs. This finding does not support the common perception that catch-up longitudinal studies limit reliability and validity because they preclude optimal measurement procedures. Although preliminary, the findings suggest this design may be a reasonable (and perhaps economically efficient) alternative to a true prospective design, providing raters are blind (Douglas, Ogloff, & Hart, 2003).

**Clinical Implications**

Findings from the current study offer support for the argument that until

empirical research indicates otherwise, professionals in the ID field are justified in using the VRAG and/or HCR-20 to assess risk of aggression (Camilleri & Quinsey, 2011; Pouls & Jeandarme, 2015) with a good level of accuracy. Dynamic scales are in the early stages of development (currently with poor psychometric properties). As research develops in the ID field and dynamic scales are developed with established psychometric properties, a future comparison is needed to indicate whether or not well-developed actuarial, SPJ and dynamic measures differ in their predictive ability.

Whilst there is extensive research comparing actuarial and SPJ approaches and staunch advocates for each method (Hanson & Morton-Bourgon, 2009; Hart & Cooke, 2013), several authors advocate a convergent approach that focuses on risk formulation (Boer, 2004; Singer et al. 2013). Using this approach, Singer and colleagues (2013) recommend assessors use a variety of measures that "converge" on the target behaviour to establish the pertinent risk issues and the appropriate level to intervene and manage risk. This would seem a sensible solution to ensure that pertinent case specific factors are accommodated in a comprehensive risk assessment. However, a limitation of this approach is the potential increased burden on busy clinicians. In recent surveys, clinicians indicated that the required training, administration and interpretation of SPJs such as HCR-20 is time-consuming and resource intensive (Green, Caroll & Brett, 2010; Nicholls, Petersen, & Pritchard, 2016). Understanding *how* to make the process of risk prediction more efficient, in terms of time and cost of training, ease of administration, and service user involvement are important to consider as tools are developed/refined.

The inclusion of studies using the PCL-R (Hare, 2003) as a measure of risk is worthy of note for several reasons. The PCL-R was designed to measure the clinical concept of psychopathy, not to assess risk of violence, general offending (Hare, 2006)

or treatment outcome. Therefore, the PCL-R should not be used within research or clinical practice to assess risk. Use of the measure for risk assessment purposes is based on the assumption that there is an inherent link between psychopathy and violence, which contributes to or increases the presence of risk. This assumption and the use of the PCL-R as a risk assessment tool is much debated within the mainstream literature. Furthermore, the validity and reliability of the construct of psychopathy as measured by the PCL-R is also widely contested. Authors argue that Hare's conceptualization of psychopathy is tautological (Ellaerd, 1988) and subjective. Other studies have found that the evaluators' personality can bias the judgments he or she make regarding whether an individual meets the criteria for a psychopathy label (Miller, Rufino, Boccaccini, Jackson & Murrie, 2011).

The concept of psychopathy raises pertinent clinical and ethical concerns. Receiving a diagnosis of 'psychopath' is stigmatising and commonly leads to the assumption that the individual is untreatable (Gendreau, Goggin, & Smith, 2002). Adults who are diagnosed with a severe personality disorder are likely to be detained in secure hospitals in the UK under the Mental Health Act (1983). Attracting a label of 'psychopath' is particularly harmful for adults with ID who are already at increased risk of stigmatisation, marginalization, and restrictions on their lives by virtue of their disability.

The present study was the first attempt to quantify the effectiveness of risk assessment measures for predicting aggression in adults with ID. A particular strength of this study was the inclusion of prospective studies only, which provided more robust evidence than retrospective studies (Hanson, 2009). This is in line with the epidemiological definition of risk as something that takes place before the outcome (Kraemer et al., 1997). The further strength of this study was the consideration of

quality of included studies (using the CASP tool) in addition to the AUCs weighted by sample size, when drawing conclusions regarding the effectiveness of the three types of measures and study design.

Despite the promising findings for the performance of risk assessments for predicting risk of aggression with individuals with ID, the present study contained only 14 studies. Primarily, this is because compared with the general offender literature, the research in this area is limited. A further limitation of the present study was the absence of inter-rater reliability at the study identification and quality assessment rating stages.

**Future research**

It is proposed that the relationship between dynamic risk factors and offending behaviour is worthy of continued research attention in ID populations. Findings from mainstream offending literature demonstrate a well-established evidence base for dynamic approaches to assessing risk in this population, a pattern that is starting to emerge in the ID field. To date, where primary research has directly compared the two types of measures, it has concluded that dynamic risk variables may be as good as or better than static variables in predicting violent and sexual incidents in offenders with ID (Blacker, et al. 2010; Lindsay et al. 2008; Lofthouse et al., 2013).

Future studies should include broader search terms and replicate the analysis for sexual and general offending behaviours to explore whether or not the pattern of findings from the current study are replicated with other types of aggression. Future research can measure how individual assessments perform across gender, ethnic group and level of ID. The research field can also move on from comparing

instruments with one another to understand how far into the future prediction is optimal with different measures.

In the current review, the majority of studies were conducted within secure settings. Given that the environment and infrastructure in secure services differs significantly to community settings, it would seem logical to assume that there are significant differences in predicting aggression in these settings. Indeed, in recent studies by Lofthouse and colleagues (Lofthouse et al., 2014b; Lofthouse, 2016) dynamic risk measures, when used in a community sample, predicted aggression with greater accuracy over a one-month period, whereas, in secure settings, optimal prediction occurred over three months. These findings would suggest that important differences occur in predicting aggression in restricted settings compared to the community or following discharge. Reliance on containment strategies in secure services to reduce and manage aggression and a more controlling environment are potential explanations for the difference. In light of the introduction of policies internationally enabling more individuals to live in the community, further research in this area is warranted.

In summary, the current study was a first endeavor to synthesise evidence from prospective studies on the prediction of aggression in individuals with ID. The studies included in this review demonstrate that existing risk assessment methods significantly predict the risk for aggression among adults with ID, with no type of instrument outperforming the other at this stage. These findings help clinicians make informed, evidence based decisions when selecting measures for assessing risk for adults with ID. It is recommended that a new meta-analysis is conducted when dynamic measures for this population reach the same level of methodological quality as existing actuarial and SPJ methods.

**References**

Ali, A., Hall, I., Blickwedel, J., & Hassiotis, A. (2015). Behavioural and cognitive- behavioural interventions for outwardly- directed aggressive behaviour in people with intellectual disabilities. *The Cochrane Library*.

Andrews, D. A. & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis.

Beech, A. R., & Ward, T. (2004). The integration of etiology and risk in sexual offenders: A theoretical framework. *Aggression and Violent Behavior*, **10**, 31-63.

Blacker, J., Beech, A. R., Wilcox, D. T., & Boer, D. P. (2010). The assessment of dynamic risk and recidivism in a sample of special needs sexual offenders. *Psychology, Crime and Law*, *17*, 75-92.

Boer, D. P., Tough, S., & Haaven, J. (2004). Assessment of risk manageability of intellectual disabled sex offenders. *Journal of Applied Research in Intellectual Disabilities, 17*, 275-283.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk- 20: Professional guidelines for assessing risk of sexual violence*. Burnaby, BC: The Mental Health, Law & Policy Institute.

Borum, R., Bartel, P., & Forth, A. (2006). Structured Assessment of Violence Risk in Youth (SAVRY). Lutz, Florida: Psychological Assessment Resources.

Borthwick-Duffy, S. A. (1994). Epidemiology and prevalence of psychopathology in people with mental-retardation. *Journal of Consulting and Clinical Psychology, 62*(1).

Camilleri, J. A., & Quinsey, V. L. (2011). Appraising the risk of sexual and violent recidivism among intellectually disabled offenders. *Psychology Crime & Law, 17*(1), 59-74.

Campbell, M. A., French, S., & Gendreau, P. (2009). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism for adult offenders*. Ottawa, ON: Public Safety Canada.

CASP. (2013). Critical Appraisal Skills Programme Cohort Study Checklist. In *http://www.casp-uk.net/wp-content/uploads/2011/11/CASP-Cohort-Study-Checklist-31.05.13.pdf* (Vol. 2013).

Cooper, S. A., Smiley, E., Jackson, A., Finlayson, J., Allan, L., Mantry, D., & Morrison, J. (2009). Adults with intellectual disabilities: prevalence, incidence and remission of aggressive behaviour and related factors. *Journal of Intellectual Disability Research, 53*(3), 217-232.

Craig, L. A., Beech, A. R, Cortini, F. (2013). What works in assessing risk in sexual and violent offenders, in Craig, L. A., Dixon, L. & Gannon, T.A. (Eds.) *What Works in Offender Rehabilitation: An Evidence-Based Approach to Assessment and Treatment,* Chichester: Wiley-Blackwell.

Crocker, A. G., Mercier, C., Lachapelle, Y., Brunet, A., Morin, D., & Roy, M. E. (2006). Prevalence and types of aggressive behaviour among adults with intellectual disabilities. *Journal of Intellectual Disability Research, 50*(9), 652–661.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk of violence – User guide*. Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.

Douglas, K. S., Ogloff, J. R., & Hart, S. D. (2003). Evaluation of a model of violence

risk assessment among forensic psychiatric patients. *Psychiatric Services: A Journal of the American Psychiatric Association, 54*(10), 1372-1379.

Dolan, M., & Doyle, M. (2007). Psychopathy: Diagnosis and implications for treatment. *Psychiatry*, *6*(10), 404-408.

Drieschner, K. H., Marrozos, I., & Regenboog, M. (2013). Prevalence and risk factors of inpatient aggression by adults with intellectual disabilities and severe challenging behaviour: A long-term prospective study in two Dutch treatment facilities. *Research in Developmental Disabilities*, *34*(8), 2407-2418.

Drieschner, K. H., & Hesper, B. L. (2008). Dynamic risk outcome scales. Boschoord, NL: Trajectum.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. British Medical Journal, 315, 629–634.

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). The prediction of violence and antisocial behaviour: A systematic review and meta-analysis of the utility of risk assessment instruments in 73 samples involving 24,827 individuals. *British Medical Journal*, *345*.

Fitzgerald, S., Gray, N. S., Alexander, R. T., Bagshaw, R., Chesterman, P., Huckle, P., . . . Snowden, R. J. (2013). Predicting institutional violence in offenders with intellectual disabilities: The predictive efficacy of the VRAG and the HCR-20. *Journal of Applied Research in Intellectual Disabilities, 26*(5), 384-393.

Fitzgerald, S., Gray, N., Taylor, J., Snowden, R., Fitzgerald, S., Gray, N. S., . . . Snowden, R. J. (2011). Risk factors for recidivism in offenders with intellectual disabilities. *Psychology, Crime & Law*, *17*(1), 43-58.

Gray, N. S., Fitzgerald, S., Taylor, J., MacCulloch, M. J., & Snowden, R. J. (2007).

Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL-SV, & the HCR-20. *Psychological Assessment, 19*(4), 474-479.

Gray ,N. S, Taylor, J., & Snowden, R. J. (2011). Predicting violence using structured professional judgment in patients with different mental and behavioral disorders. *Psychiatry Research, 187*(1-2), 248-253

Green, B., Caroll, A., & Brett, A. (2010). Structured risk assessment in community forensic mental health practice. *Australasian Psychiatry*, *18,* 538–541.

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology-Psychologie Canadienne, 50*(3), 172-182.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*(1), 1-21.

Hare R. D. (2003) *The Hare Psychopathy Checklist-Revised* (2nd ed.). Toronto, Ontario: Multi-Health Systems.

Hart, S. D., & Cooke, D. J. (2013). Another look at the precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences & the Law*, *31*(1), 81-102.

Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. In P. Sturmey & M. McMurran (Eds.), *Forensic case formulation* (pp. 83–106). Chichester, UK: Wiley-Blackwell.

Hastings, R. P. (2002). Do challenging behaviors affect staff psychological well-being? Issues of causality and mechanism. *American Journal on Mental Retardation, 107*(6), 455-467.

Hastings, R. P., & Brown, T. (2002). Coping strategies and the impact of challenging behaviors on special educators' burnout. *Mental Retardation, 40*(2), 148-156.

Hensel, J. M., Lunsky, Y., & Dewa, C. S. (2012). Exposure to client aggression and burnout among community staff who support adults with intellectual disabilities in Ontario. Canada. *Journal of Intellectual Disability Research, 56(9),* 910–915.

Hurducas, C., Singh, J., de Ruiter, C., & Petrila, J. (2016). Violence Risk Assessment Tools: A Systematic Review of Surveys. In J. P. Singh, S. Bjørkly, & S. Fazel (Eds.), *International Perspectives on Violence Risk Assessment* (pp. 76-100). United Kingdom: Oxford University Press.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560.

Hockenhull, J. (n.d.). The structured assessment of violence risk in adults with intellectual disability: A systematic review. Liverpool Reviews and Implementations Group

Inett, A., Wright, G., Roberts, L., & Sheeran, A. (2014). Predictive validity of the START with intellectually disabled offenders. *Journal of Forensic Practice, 16*(1), 78-88.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupler D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54,* 337–343.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). Spousal Assault Risk Assessment guide user's manual. *Toronto, Ontario, Canada: Multi-Health Systems*.

Lim, S., & Howard, R. (1998). Antecedents of sexual and non-sexual aggression in young Singaporean men. *Personality and Individual Differences*, *25*(6), 1163-1182.

Lindsay, W. R., Hogue, T. E., Taylor, J. L., Steptoe, L., Mooney, P., O'Brien, G., . . . Smith, A. H. W. (2008). Risk assessment in offenders with intellectual disability. A comparison across three levels of security. *International Journal of Offender Therapy and Comparative Criminology, 52*(1), 90-111.

Lofthouse, R. E. (2016). *Assessing and managing risk with adults with intellectual disabilities.* (Unpublished doctoral dissertation). University of Liverpool, UK.

Lofthouse, R. E., Lindsay, W. R., Totsika, V., Hastings, R. P., Boer, D. P., & Haaven, J. L. (2013). Prospective dynamic assessment of risk of sexual reoffending in individuals with an intellectual disability and a history of sexual offending behaviour. *Journal of Applied Research in Intellectual Disabilities*, *26*(5), 394-403.

Lofthouse, R. E., Lindsay, W. R., Totsika, V., Hastings, R. P., & Roberts, D. (2014b). Dynamic risk and violence in individuals with an intellectual disability: tool development and initial validation. *The Journal of Forensic Psychiatry & Psychology*, *25*(3), 288-306.

Lofthouse, R. E., Totsika, V., Hastings, R. P., Lindsay, W. R., Hogue, T. E., & Taylor, J. L. (2014a). How do static and dynamic risk factors work together to predict violent behaviour among offenders with an intellectual disability?. *Journal of Intellectual Disability Research*, *58*(2), 125-133.

Mills, J. F. (2017). Violence risk assessment: A brief review, current issues, and future directions. *Canadian Psychology/Psychologie canadienne*, *58*(1), 40.

MedCalc Statistical Software version 16.4.3 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2016).

Mills, S., & Rose, J. (2011). The relationship between challenging behaviour, burnout and cognitive variables in staff working with people who have intellectual disabilities. *Journal of Intellectual Disability Research, 55(*9), 844–857.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269.

Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, …, & Banks, S. (2001). *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*. New York: Oxford University Press.

Morrissey, C., Hogue, T., Mooney, P., Allen, C., Johnston, S., Hollin, C.,…Taylor, J. L. (2007). The predictive validity of the PCL-R in offenders with ID in a high secure hospital setting: *Institutional aggression. Journal of Forensic Psychiatry & Psychology, 18,* 1–15.

Morrissey, C., Hogue, T., Mooney, P., Lindsay, W. R., Steptoe, L., Taylor, J. & Johnston, S. (2005). Applicability, reliability, and validity of the psychopathy checklist-revised in offenders with intellectual disabilities: Some initial findings. *International Journal of Forensic Mental Health, 4,* 207–20.

O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014). Differential predictive validity of the Historical, Clinical and Risk Management Scales (HCR–20) for inpatient aggression. *Psychiatry Research, 220*(1), 669-678.

Pouls, C., & Jeandarme, I. (2014). Psychopathy in Offenders with Intellectual Disabilities: A Comparison of the PCL-R and PCL: SV. *International Journal of Forensic Mental Health, 13*(3), 207-216.

Pouls, C., & Jeandarme, I. (2015). Risk assessment and risk management in offenders with intellectual disabilities: Are we there yet? *Journal of Mental Health Research in Intellectual Disabilities, 8*(3-4), 213-236.

Prout, T. H., & Strohmer, D. C. (1991). *Emotional Problems Scales*. Lutz, FL: Psychological Assessment Resources.

Quinsey, V. L. (2004). Risk assessment and management in community settings. In W. R. Lindsay, J. L. Taylor, & P. Sturmey (Eds.), *Offenders with developmental disabilities* (pp. 131-141). Chichester, UK: John Wiley & Sons Ltd.

Quinsey, V. L., Book, A. B., & Skilling, T. A. (2004). A Follow-up of deinstitutionalized men with intellectual disabilities and histories of antisocial behaviour. *Journal of Applied Research in Intellectual Disabilities, 17*, 243-253.

Quinsey, V. L., Coleman, G., Jones, B., & Altrows, I. F. (1997). Proximal antecedents of eloping and reoffending among supervised mentally disordered offenders. *Journal of Interpersonal Violence*, *12*(6), 794-813.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and Managing Risk (1st edition).* Washington, DC; American Psychological Association.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk (2nd ed.).* Washington, DC, US: American Psychological Association.

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*(5), 615-620.

Schoonjans, F., Zalata, A., Depuydt, C. E., Comhaire, F. H. (1995). MedCalc: A new computer program for medical statistics. *Computer Methods and Programs in Biomedicine, 48*(3), 257.

Singer, J. C., Boer, D. P., & Rettenberger, M. A Convergent Approach to Sex Offender Risk Assessment. In K. Harrison & B Rainey (Eds.), *The Wiley-Blackwell Handbook of Legal and Ethical Aspects of Sex Offender Treatment and Management* (pp. 339-355). Chichester, UK: John Wiley & Sons.

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior, 37*(9), 965-988.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical psychology review*, *31*(3), 499-513.

Singh, J. P., Grann, M., & Fazel, S. (2013). Authorship bias in violence risk assessment? A systematic review and meta-analysis. *PLoS ONE 8*(9): e72484.

Steptoe, L. R., Lindsay, W. R., Murphy, L., & Young, S. J. (2008). Construct validity, reliability and predictive validity of the dynamic risk assessment and management system (DRAMS) in offenders with intellectual disability. *Legal and Criminological Psychology, 13*, 309-321.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). Manual for the Short-Term Assessment of Risk and Treatability (START), Version 1.0. St. Joseph's Healthcare Hamilton: Ontario, Canada–Forensic Psychiatric Services Commission: Port Coquitlam. *British Columbia, Canada*.

Wheeler, J. R., Clare, I. C. H., Holland, A. J. (2014). What can social and environmental factors tell us about the risk of offending by people with intellectual disabilities? *Psychology, Crime & Law, 20*(7), 635-658.

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk scale - A treatment-friendly violence risk assessment tool. *Psychology Public Policy and Law, 12*(3), 279-309.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*(5), 740-767.

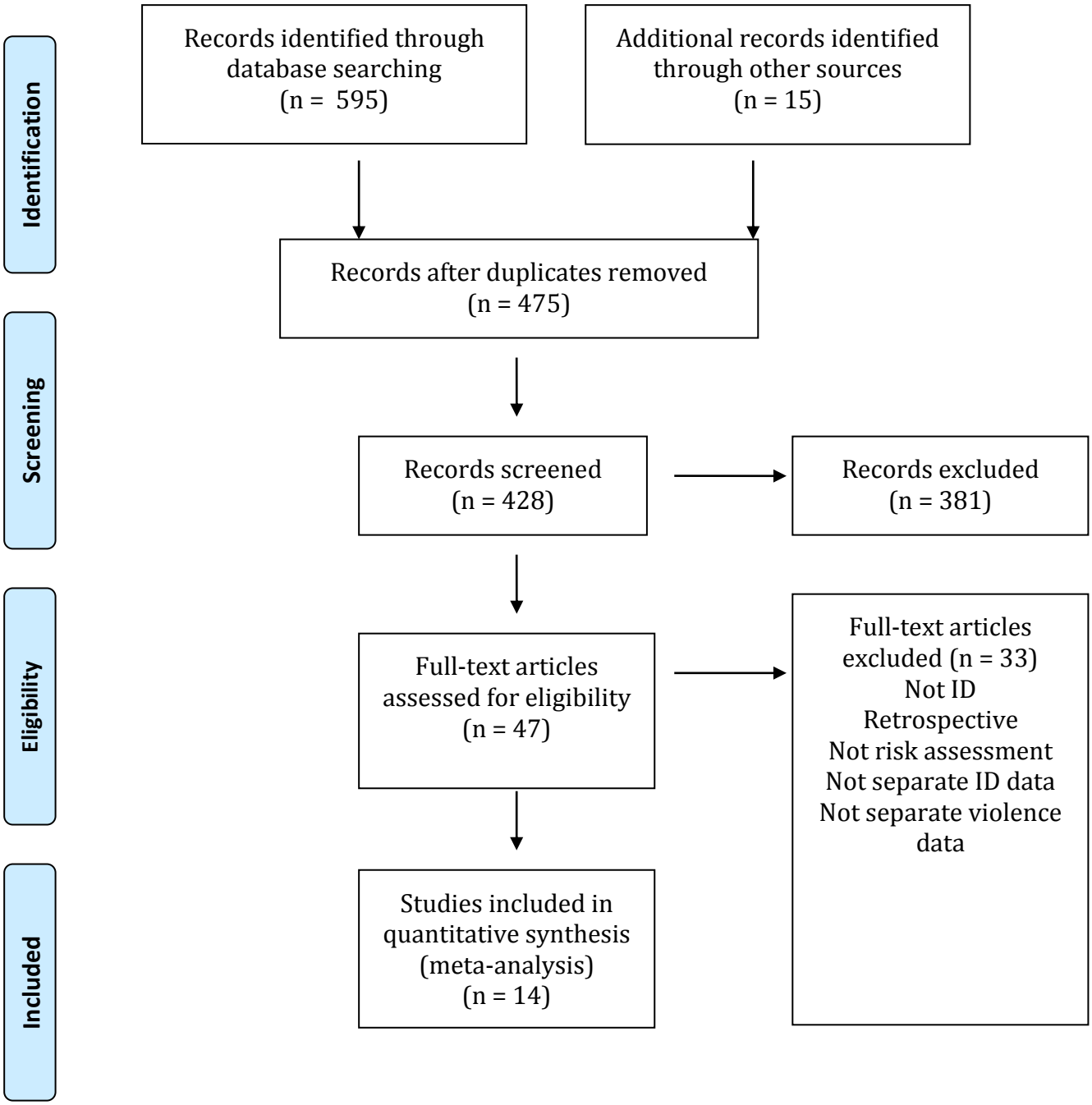


*Figure 1.* Results of a systematic search conducted to assess the effectiveness of available risk measures for predicting aggression among adults with ID

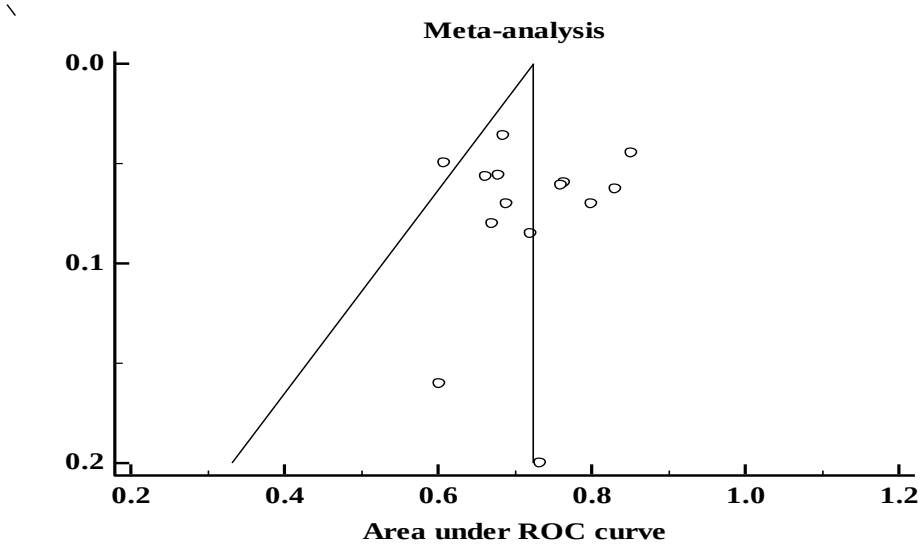


*Figure 2.* Funnel plot of the effect size against the standard error for 14 studies included in the meta-analysis. The vertical line represents the summary effect size.

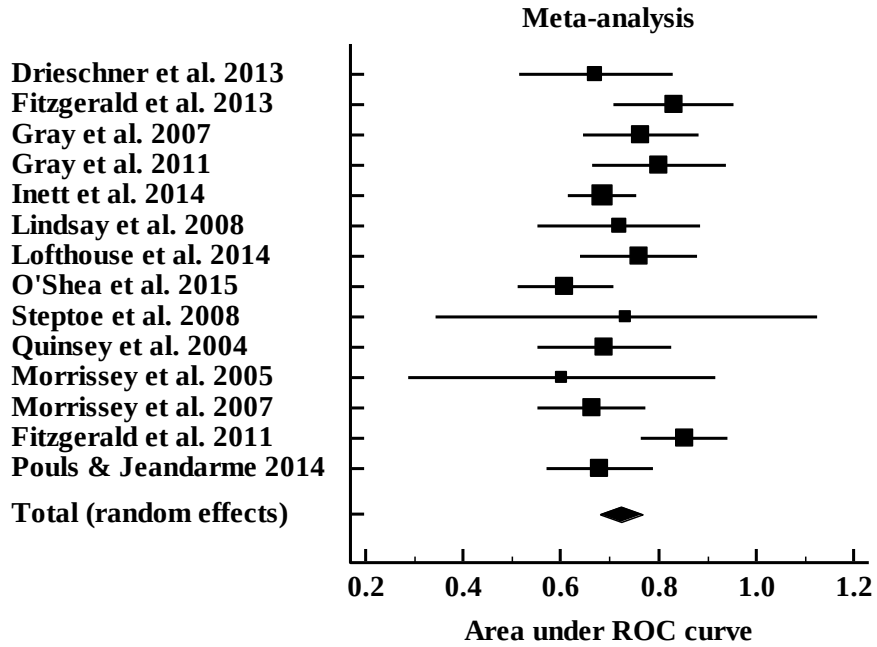


*Figure 3.* A forrest plot of effect size and 95% confidence intervals (CIs) for the 14 studies included in the meta-analysis.

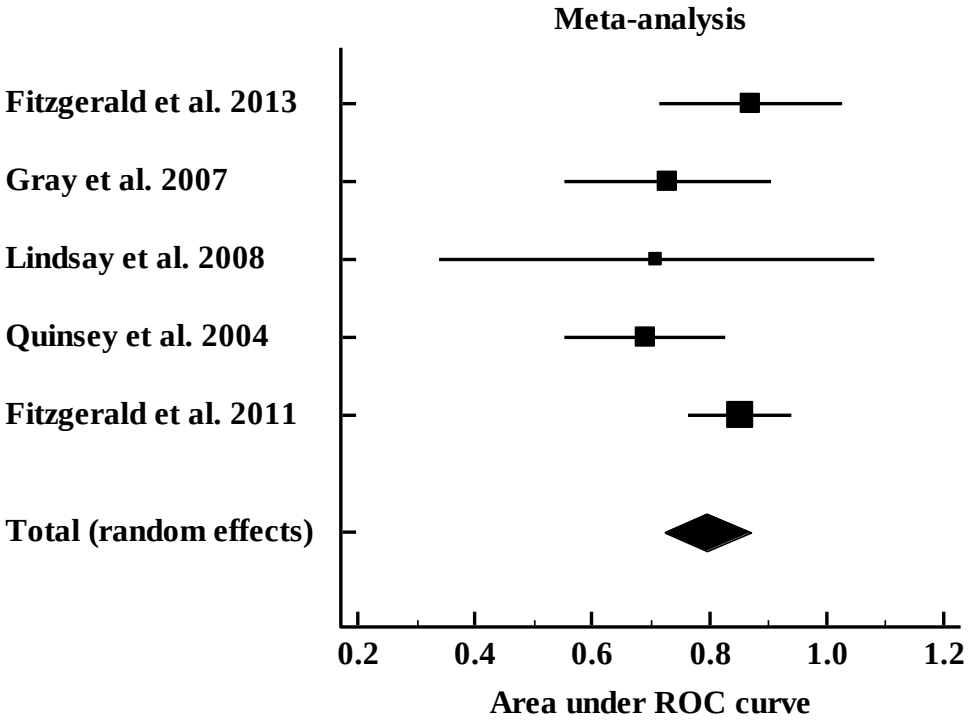


*Figure 4.* A forrest plot of effect size and 95% confidence intervals (CIs) for the five studies included actuarial measures.

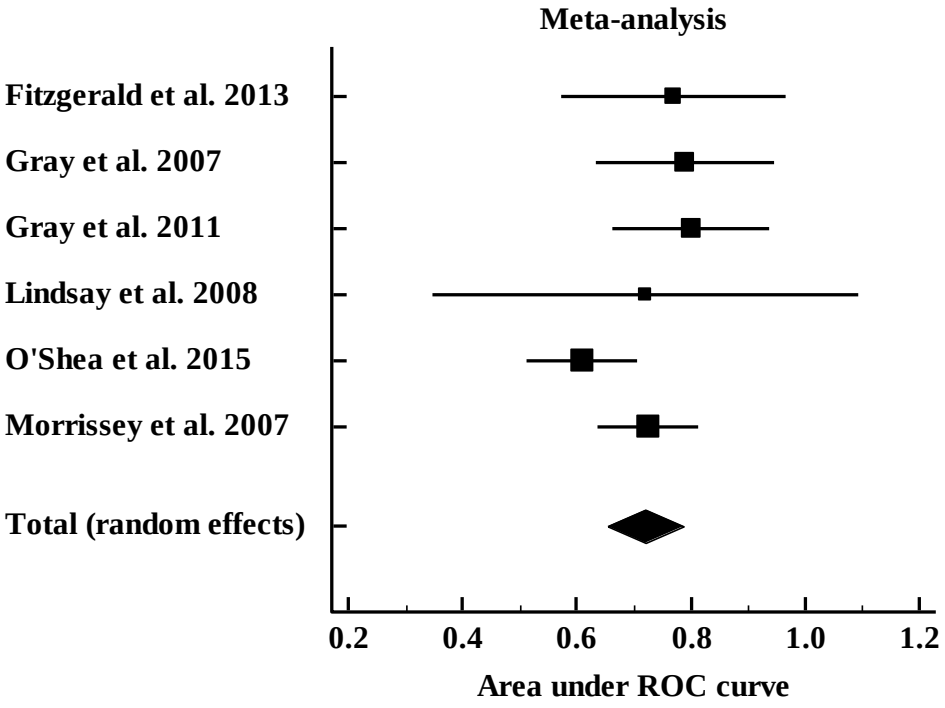

*Figure 5.* A forrest plot of effect size and 95% confidence intervals (CIs) for the six studies included SPJ measures.

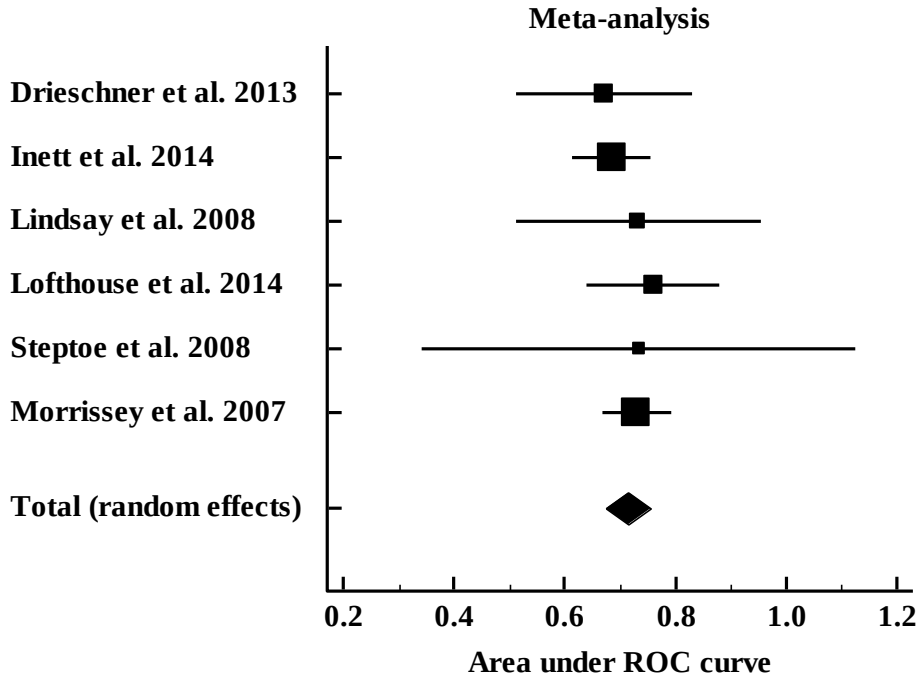


*Figure 6.* A forrest plot of effect size and 95% confidence intervals (CIs) for the six studies included dynamic measures.

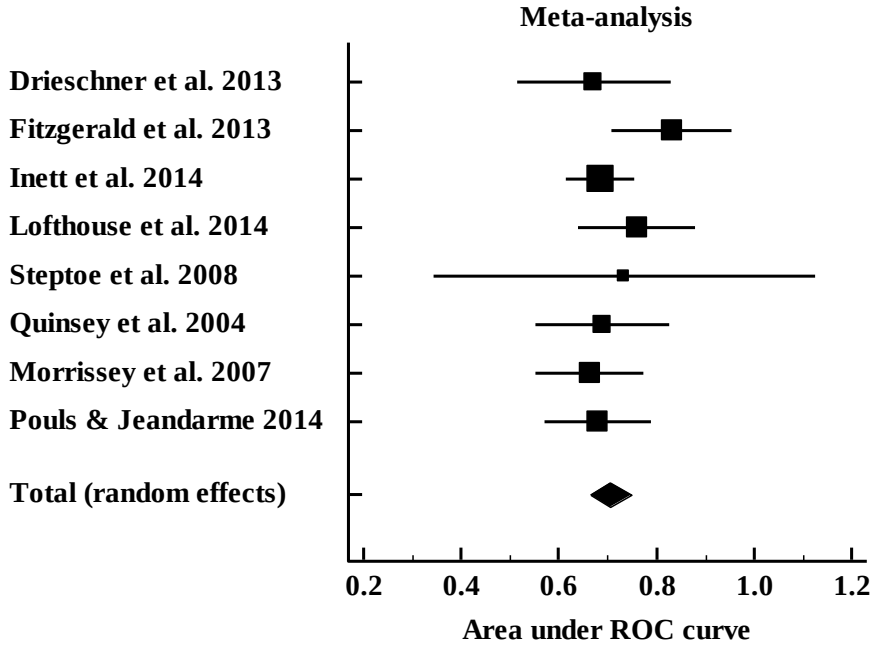

*Figure 7.* A forrest plot of effect size and 95% confidence intervals (CIs) for the eight studies including a prospective design.

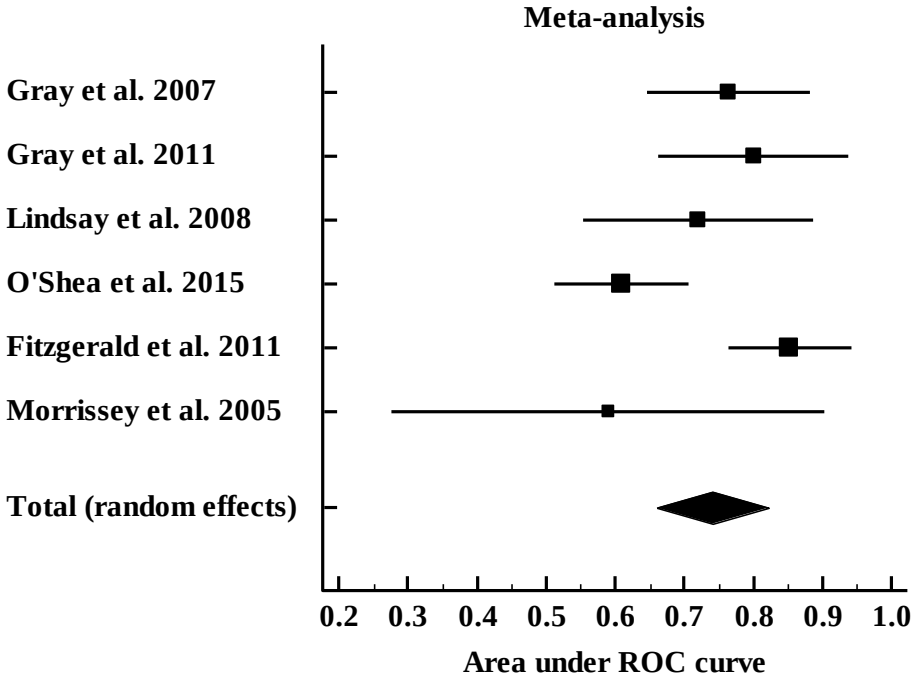


*Figure 8.* A forrest plot of effect size and 95% confidence intervals (CIs) for the six studies including a catch-up longitudinal design.

Table 1
*Characteristics of studies identified for inclusion (n=14)*

| Author/year | Country | Design | N | Age (mean) | ID definition | Gender | Setting | Measure | Measure type | Quality Assessment |
|---|---|---|---|---|---|---|---|---|---|---|
| Quinsey et al. (2004) | Canada | Prospective | 58 | 40.61 years (SD . 10.59, n . 57). | Mental retardation | 58 m | Residential institution | VRAG PIC PRFS | Act Dynamic | 12 |
| Gray, et al. (2007) | UK | Catch-up longitudinal | 145 | 30.8 | ICD 10 MR (F70—F79). 121 mild, 18 moderate, 5 severe, 1 unspecified | 118 m 27 f | Discharged from medium secure psychiatric unit | VRAG, HCR-20, PCL-SV | Act SPJ | 13 |
| Morrissey et al. (2005) | UK | Catch-up longitudinal | 203 | 37 | Mean IQ 66 | 203 m | Forensic ID high security hospital | PCLR EPS-BRS | Act Dynamic | 14 |
| Morrissey et al. (2007) | UK | Prospective | 73 | 38 | Mean IQ 66.6 | 73 m | Forensic ID high security hospital | PCLR EPS-BRS | Act Dynamic | 17 |
| Steptoe et al. (2008) | UK | Prospective | 23 | 38.4 | Mean IQ 64 | 23 m | Forensic ID high security hospital | DRAMS | Dynamic | 14 |
| Lindsay et al. (2008) | UK | Catch-up longitudinal | 212 | High 38.7, med/low 39.0, community 34.3 | High Mean IQ 66.6, Med/low Mean IQ 66.7, community Mean IQ 64.7 | 212 m | High, med/low, community | EPS VRAG, HCR20 SDRS | Dynamic Act, SPJ | 14 |
| Gray et al. (2011) | UK | Catch-up longitudinal | 115 | 37.7 | ICD-10 (F70-79) mental retardation | U | Discharged medium secure psychiatric units | HCR20 | SPJ | 12 |
| Fitzgerald et al. (2011) | UK | Catch-up longitudinal | 85 | 31.54 | Unspecified | U | Discharged from medium secure units | OGRS | Act | 15 |
| Drieschner et | Netherland | Prospective | 218 | 33.8 | Mean IQ 70.3 | 86.4% m | Residential. | Dynamic | Dynamic | 15 |

| Author/year | Country | Design | N | Age (mean) | ID definition | Gender | Setting | Measure | Measure type | Quality Assessment |
|---|---|---|---|---|---|---|---|---|---|---|
| al. (2013) | s | | | | | | Forensic & non forensic | risk outcome Scale (DROS). | | |
| Fitzgerald et al. (2013) | UK | Prospective | 25 | 29.77 | Mean IQ 64.59 | 23 m 2 f | Medium secure unit | HCR20 VRAG | SPJ Act | 15 |
| Inett et al. (2014) | UK | Prospective | 27 | 39 | Learning disability | U | Low secure | START | Dynamic | 12 |
| Lofthouse et al. (2014) | UK | Prospective | 64 | 41.9 | Administratively defined | 45 m 19 f | Forensic unit, rehabilitation, acute mental health, residential service, hospital setting | CuRV | Dynamic | 12 |
| Pouls & Jeandarme (2014) | Belguim | Prospective | 52 | 40 | Mean IQ 57 | 52 m | Forensic unit or prison | PCLR PCL SV | Act | 14 |
| O'shea et al. (2015) | UK | Catch-up longitudinal | 109 | 32 | ICD-10 MR | 70 m 39 f | Inpatient mental health | HCR20 | SPJ | 11 |

VRAG = Violence Risk Appraisal Guide; PIC = SPJ = structured professional judgement; Act = actuarial; U = unspecified ; m= male; f = female; PIC = Problem Identification Checklist (Quinsey et al., 1997); PRFS = Proximal Risk Factor Scale, (Quinsey et al., 1997); DRAMS = Dynamic Risk Assessment and Management System, (Lindsay et al.., 2004); DROS = Dynamic Risk Outcome Scale, (Drieschner & Hesper, 2008); START = Short-Term Assessment of Risk and Treatability, (Webster, Martin, Brink, Nicholls & Middleton, 2004), SDRS = Short Dynamic Risk Scale, (Quinsey 2004); CuRV = Current Risk of Violence, (Lofthouse, Lindsay, Totsika, Hastings, & Roberts, 2014).

Table 2

*Effect size and 95% confidence interval for type of risk assessment*

| *Risk assessment type* | *N* | *Area Under Curve* | *95 % Confidence Interval* |
|---|---|---|---|
| Actuarial | 5 | 0.796 | 0.723, 0.869 |
| SPJ | 6 | 0.721 | 0.654, 0.788 |
| Dynamic | 6 | 0.714 | 0.673, 0.756 |

Table 3

*Effect size and 95% confidence interval for study design*

| *Study design* | *N* | *Area Under Curve* | *95 % Confidence Interval* |
|---|---|---|---|
| Prospective | 8 | 0.706 | 0.655, 0.747 |
| Catch-up longitudinal | 6 | 0.741 | 0.661, 0.822 |