Estimation of the Minimal Clinically Important Difference on the Aberrant Behaviour Checklist –Irritability (ABC-I) for people with intellectual disabilities who display aggressive challenging behaviour: a triangulated approach

Key words: intellectual disabilities, aggressive challenging behaviour, MCID, RCT, clinical trials

What this paper adds

This is the first study to investigate the role of the minimal clinically important difference in the conduct of clinical trials in the field of intellectual disabilities. It uses the example of a widely used measure of aggressive challenging behaviour in establishing a clinically meaningful difference that can be used in the apriori calculation of a trial sample size.

Abstract

Background

The minimal clinically important difference (MCID) is relevant in the estimation of improvement in a patient outcome.

Aim

To determine the MCID on the Aberrant Behaviour Checklist–Irritability (ABC-I), widely used to measure the effects of intervention for aggressive challenging behaviour in people with intellectual disabilities.

Method and Procedures

We utilised distribution and anchor based methods to estimate the ABC-I MCID. We extracted data from 15 randomised controlled trials (RCTs) for meta-analysis. We conducted three online workshops with family carers and professionals to consider meaningful change in case vignettes of increasing severity of aggressive challenging behaviour.

Outcomes and Results

We did not find overlap in the range of values between the two approaches. The meta-analysis indicated a range of MCID on the ABC-I (0.05, 4.94) whilst the anchor-based estimation indicated a larger change (6.6, 16.6).

Conclusions and Implications

The MCID is essential in interpreting the results from intervention studies. The present work was undertaken as part of a wider programme on the development and testing of a psychosocial intervention for aggressive challenging behaviour, and it is of interest to researchers in justifying how they choose and determine the MCID on the outcome of interest.

Word count: 195

1.1 Introduction

One in six people has a developmental disability including intellectual disability, autism and other conditions often co-occurring in the same person (Zablotsky et al., 2019; McGuire et al., 2019). They are more vulnerable to develop a mental health condition, have poorer health and decreased life expectancy (Glover et al., 2017; Hughes-McCormack et al., 2021). Multiple sources have identified high personal, care and societal costs associated with intellectual disabilities internationally (Fujiura et al., 2018, Lunsky et al., 2019, Arora et al., 2020). One of the most pressing areas of need is the management of aggressive challenging behaviour in people with intellectual disabilities. That includes threatening or intimidating behaviours or non-compliance and against social norms and aggression towards others or property. Therefore, development and provision of clinically and cost-effective interventions for aggressive challenging behaviour is a public health priority.

The concept of the minimal clinically important difference (MCID) was first introduced in 1984 to address the reporting of findings in psychotherapy trials. MCID is defined as: 1. “the smallest difference in outcome which study patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management” (Jacobson et al, 1984); 2. “the smallest change that is important to patients” (Copay et al, 2007); 3. “the smallest effect size that would lead health professionals to recommend a therapy to their patients” (Hays & Woolley, 2000). MCID is different from the concept of reliable change which is a function of the psychometric properties of a given outcome measure. Often, the MCID reflects the extent of improvement from baseline with higher baseline severity accounting for greater perception of benefit (Button et al., 2015).

Several approaches are available to estimate the MCID on an outcome measure. *Distribution-based* methods estimate the MCID from existing datasets using group differences through time or between groups at a specific time point in relation to the standard deviation at baseline (Hays & Woolley, 2000; Wright et al, 2012).

To calculate sample sizes for primary outcome measures without a pre-determined MCID, researchers conventionally use a moderate effect size of 0.5xStandard Deviation (SD) (Norman et al, 2003), though a range of effect sizes maybe considered from as low as 0.2 standard deviation (SD), to large of 0.80SD (Cohen 1988).

In *anchor-based* approaches, variation in an outcome reported by patients is compared to a benchmark or 'anchor' such as clinician or patient opinion (Hays & Woolley, 2000; Wright et al, 2012). Button et al. (2015) calculated the MCID for the Beck Depression Inventory-II (BDI-II), consistent with the global improvement rating by participants, and examined if the depression severity at baseline is associated with variations in the MCID. The authors argue that where symptom severity is significant at baseline a larger score reduction in the outcome is required to signal improvement.

Many randomised controlled trials do not routinely report MCID on the primary outcome in their sample size estimation and none has done so in the field of intellectual disabilities. This is a significant gap as clinical trials are the gold standard of intervention effectiveness and there is an ethical imperative in ensuring that they enrol enough participants to enable meaningful differences to emerge. Ultimately, the calculation of the sample size should indicate what the researchers consider to be a clinically important treatment effect.

We aimed to carry out an estimation of the MCID on ABC-I which we undertook as part of the development and testing of a new multimodal intervention for aggressive challenging behaviour in adults with intellectual disabilities. We triangulated findings from distribution and anchor based methods to arrive at what could be considered a realistic target difference to be used in apriori sample size calculations.

1.2 Method

Aberrant Behaviour Checklist – Irritability (ABC-I)

A widely used outcome measure in studies of interventions for aggressive behaviour (across the lifespan is the Aberrant Behaviour Checklist - Irritability (ABC-I; originally named "Irritability, Agitation and Crying"), which is one of the five domains of the Aberrant Behaviour Checklist – Community (ABC-C), (Aman & Singh, 2017 second edition). ABC-I consists of 15 items rated 0 (not a problem) to 3 (the problem is severe) and a total score range from 0 to 45. The ABC-I has internal consistency of 0.95 (Cronbach's alpha) and test-retest reliability of 0.98. It is deemed to be sensitive to both pharmacological and behavioural interventions. Despite its use in multiple clinical trials, there has not been yet an agreed Minimal Clinically Important Difference (MCID) on the ABC-I. The scale developers acknowledge the point of clinical change following an intervention and the research requirements of powering clinical trials adequately if using the ABC-I as primary outcome. However,they report that they have not carried out relevant studies to confirm the suggestion of a cut-off of 18 points in the ABC-I (half a standard deviation) when performing sample size calculations (ABC-2 manual, 2017, p47). These points are significant to ensure that intervention research adheres to ethical and

scientific standards by recruiting sufficient number of participants to answer the research question based on a prespecified difference between interventions.

1.1.1 Distribution-based Approach

We searched the published literature (PubMed, PsycINFO and Web of Science) until September 2020, and included papers if:

1. described a randomised controlled trial.
2. published from 2000 to September 2020 to avoid high volume of non-relevant papers,
3. the sample consisted of at least 70% of children or adults with intellectual disability with reported stratified data,
4. included any intervention (pharmacological or psychosocial/complex) and reported usable data for ABC-I as primary or secondary outcome

We also carried out forward citation searches of identified studies (Hassiotis et al., 2009; Hassiotis et al., 2011; Hassiotis et al., 2018; McCracken et al., 2002; Willner et al., 2013) and cross referenced with those listed in NICE Guideline 11 (NICE, 2015). Two researchers screened all the papers, and the final inclusion of articles was agreed with arbitration by a senior researcher (AH).

We contacted the authors of Snyder et al. (2002) to clarify the mean intervention ABC-I scores at 6-weeks, which was confirmed to be different from that included in the paper (Table 2, p 1031).

The Cochrane Risk of Bias 2 (Sterne et al., 2019) was applied to assess study quality.

Where possible we calculated means and standard deviations (SD) of the ABC-I (Higgins et al., 2020). However, we were unable to do so for Aman et al. (2002) and Carminati et al. (2016) and therefore, these data were excluded from the analyses. Studies were categorised by

1. Study population (adults only, children only, both adults and children)
2. Follow-up (short term to 3 months, medium term to 6 months and long term longer than 6 months). Where there was more than one estimate for a follow-up time, the closest to the time point of interest was selected, such that each study could have a maximum of three measures (one for each follow-up time)
3. Type of intervention (psychosocial/complex, pharmacological or both)
4. Risk of bias (low risk, some concerns and high risk)

We generated summary statistics and carried out random effects meta-analyses for each of these categories.

To use the effect size method of estimating the MCID, we determined the difference we would anticipate for a range of effect sizes, i.e., 0.5, 0.3, and 0.2 of the mean baseline intervention arm SD. We also computed change scores for the intervention arm (follow-up minus baseline) by length of follow-up and then calculated 0.5SD of the change score as a possible MCID (Wright, 2012). It was not possible to compute this for long term follow-up as there was only one such study that included baseline scores.
All data were analysed using Stata v16 (https://www.stata.com/).

1.1.2 Anchor-based Approach

We developed four clinical scenarios relating to baseline ABC-I severity bands of 11-20, 21-30, 31-40 and 41-45 taken from a previous multicentre randomised controlled trial of staff training in Positive Behaviour Support (Hassiotis et al., 2018). A statistician extracted the ABC-I scores randomly for two participants from each severity band. Only the highest scores per severity band were selected for the four case vignettes (19, 29, 39 and 45). We conducted three 90-minute online workshops with psychiatrists, healthcare professionals and family carers respectively to elicit their opinions on what they would consider to be a clinically significant change six months after treatment with a psychosocial intervention. It is common in clinical pathways in England, UK, that this amount of time may be needed from taking on a referral for treatment, followed by delivery of a psychosocial intervention and review. Although, we often find that many people with intellectual disabilities who display aggressive behaviour receive psychotropic medications, pharmacological treatments are not recommended for first line treatment use in people with intellectual disabilities unless there is imminent risk to the person or others (NICE, 2015).

Each of the workshops was attended by 1. up to 40 psychiatrists of both sexes specialising in psychiatry of intellectual disability half of whom were at Consultant grades; 2. nine female healthcare professionals from a nursing, psychology, speech and language therapy and behavioural therapy; 3. six parent or family carers of a person with intellectual disability. Most family carers worked or volunteered for charities for intellectual disability or challenging behaviour and were recruited by the Challenging Behaviour Foundation. Participants in all 3 meetings were asked, based on starting point, to rate the change in each item for it to be important in their professional or caring capacity. An anonymous online polling platform was used to record the participants' scores on each item of the ABC-I for each vignette. . Not all participants recorded a score for each item of the ABC-I.

Scores for each rater within each workshop were computed, and descriptive statistics calculated for all 15 items for a given vignette. Where a score was not given, this was considered as missing and if a range of scores was given, e.g., 0-1 or 0-2, for both of these, the responses were changed to 1 (the higher of 0-1 and the midpoint of 0-2). We calculated medians and interquartile ranges for each stakeholder group per vignette. To examine the within vignette score change, we calculated the mean difference within stakeholder groups indicating improvement, i.e. lower score (Wright et al, 2012).

1.3 Results

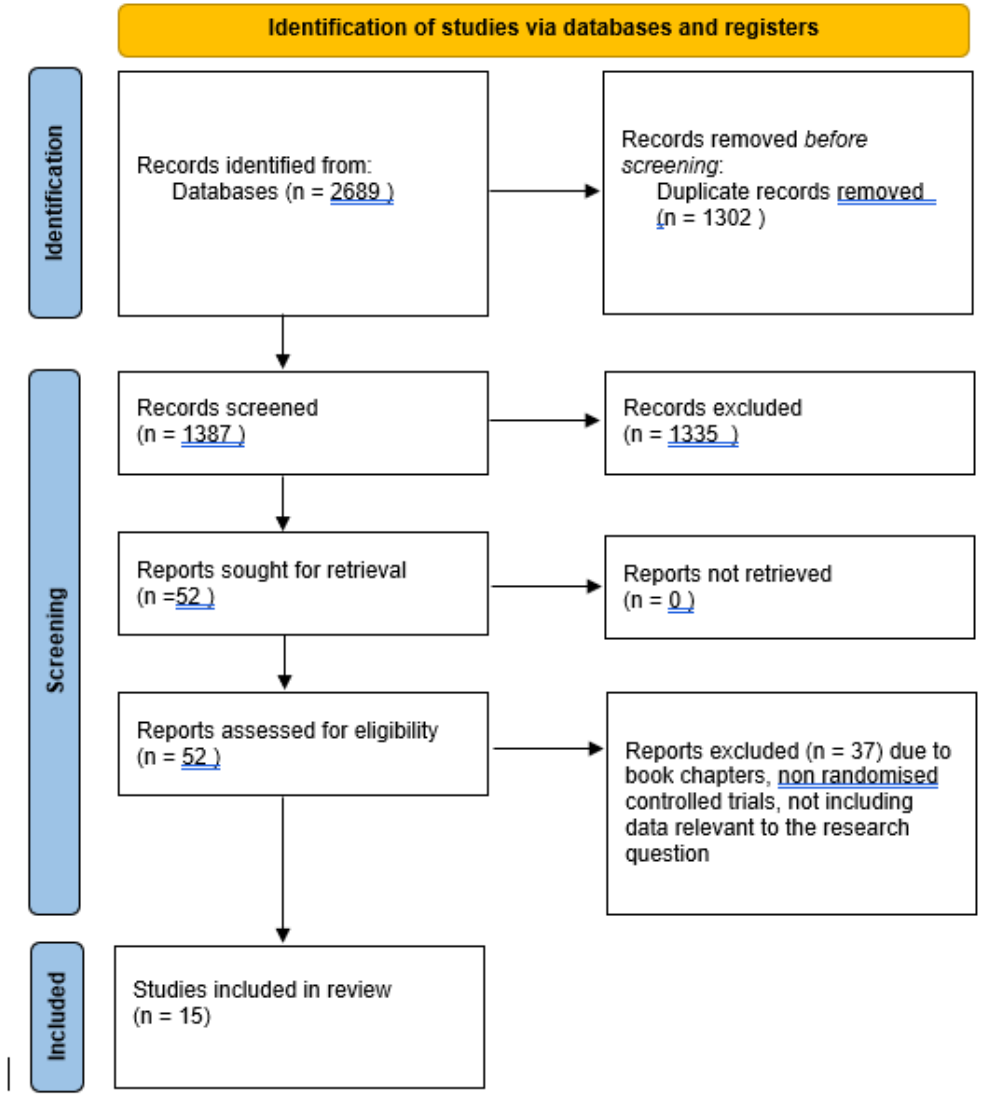
1.3.1 Distribution-based

A total of 2689 papers were identified. After de-duplication and screening of titles and abstracts, 52 full text studies were assessed for eligibility for inclusion and data extraction (Figure 1).

We finally identified 15 studies (Hassiotis et al., 2009; Hassiotis et al., 2011; Hassiotis et al., 2018; McCracken et al., 2002; Willner et al., 2013; Aman et al., 2002; Gagiano et al., 2005; Gringras et al., 2012; Hellings et al., 2005; Snyder et al., 2002; Carminati et al., 2016; Chez et al., 2020; de la Torre et al., 2020; Munesue et al., 2016; Ramerman et al., 2019) with 557 participants in the treatment and 605 in the treatment as usual/waiting list/control arms. Six studies recruited fewer than 50 participants each (Hellings et al., 2005; Munesue et al., 2016; Carminati et al., 2016; de la Torre et al., 2020; Ramerman et al., 2019; Chez et al., 2020). Four studies used the ABC-I as the primary outcome and one study as secondary outcome, whilst the remainder reported relevant ABC-I data as part of the ABC-C. There were 9 studies with short-term follow up, 6 with medium-term and 3 with long-term.

[figure 1 near here]

Figure 1: PRISMA flow diagram of studies



Ten studies reported pharmacological treatments (risperidone, valproate, venlafaxine, dextromethorphan/quinidine, melatonin and oxytocin) (Aman et al., 2002; Carminati et al., 2016; Chez et al., 2020, Gagiano et al., 2005; Gringras et al., 2012; Hellings et al., 2005; McCracken et al., 2002; Munesue et al., 2016; Ramerman et al., 2019; Snyder et al., 2002). The study by Ramerman et al. (2019) was a discontinuation study, therefore, decreasing the antipsychotic was the intervention and maintaining its administration was the control condition. Four studies were of complex interventions (positive behaviour support, applied behaviour analysis and cognitive behavioural anger management) (Hassiotis et al., 2009; Hassiotis et al., 2011; Hassiotis et al, 2018; Willner, 2013) and one combined intervention (epigallocatechin-3-gallate and cognitive training) (de la Torre et al., 2020). The mean intervention arm ABC-I score was lower than the control arm across follow-up times,

intervention types, study populations and risk of bias categorisations except for De La Torre (2020) reporting on an intervention comprising complex and pharmacological components. Details of the studies are shown in Table 1.

[table 1 near here]

Table 1: Characteristics of included studies

Author and year	Study design	Study population	Type of intervention	Follow up time points	Sample size at baseline	Risk of Bias-2 rating
Hassiotis et al. (2018)	Cluster RCT	Adults	Psychosocial – positive behaviour support	6 months, 12 months	N=245	Some concerns
Hassiotis et al. (2009)	RCT	Adults	Psychosocial – applied behaviour analysis	3 months, 6 months	N=63	Some concerns
Hassiotis et al. (2011)	RCT (naturalistic follow up)	Adults	Psychosocial – applied behaviour analysis	2 years	N=63	Some concerns
McCracken et al. (2002)	RCT	Children	Pharmacological - risperidone	8 weeks	N=101	High risk
Willner et al. (2013)	Cluster RCT	Adults	Psychosocial – cognitive behavioural anger management	4 months, 10 months	N=181 (participants with intellectual disability); N=181 (keyworkers); N=130 (home carers)	High risk
Aman et al. (2002)	RCT	Children	Pharmacological - risperidone	6 weeks	N=118	High risk
Gagiano et al. (2005)	RCT	Adults	Pharmacological - risperidone	4 weeks	N=77	High risk
Gringras et al. (2012)	RCT	Children	Pharmacological - melatonin	3 months	N=146	Some concerns
Hellings et al. (2005)	RCT	Children and adults	Pharmacological - valproate	8 weeks	N=36	High risk
Snyder et al. (2002)	RCT	Children	Pharmacological - risperidone	6 weeks	N=110	High risk
Munesue et al. (2016)	RCT crossover	Children and adults	Pharmacological - oxytocin	2, 4, 6 , 8, 10, 12, 14, 16, 20 weeks	N=29	Some concerns
Carminati et al. (2016)	RCT	Adults	Pharmacological - venlafaxine	8 weeks	N=13	High risk
de la Torre et al. (2020)	RCT	Adults	Both - epigallocatechin-3-gallate and cognitive training	3 months, 6 months	N=27	Some concerns

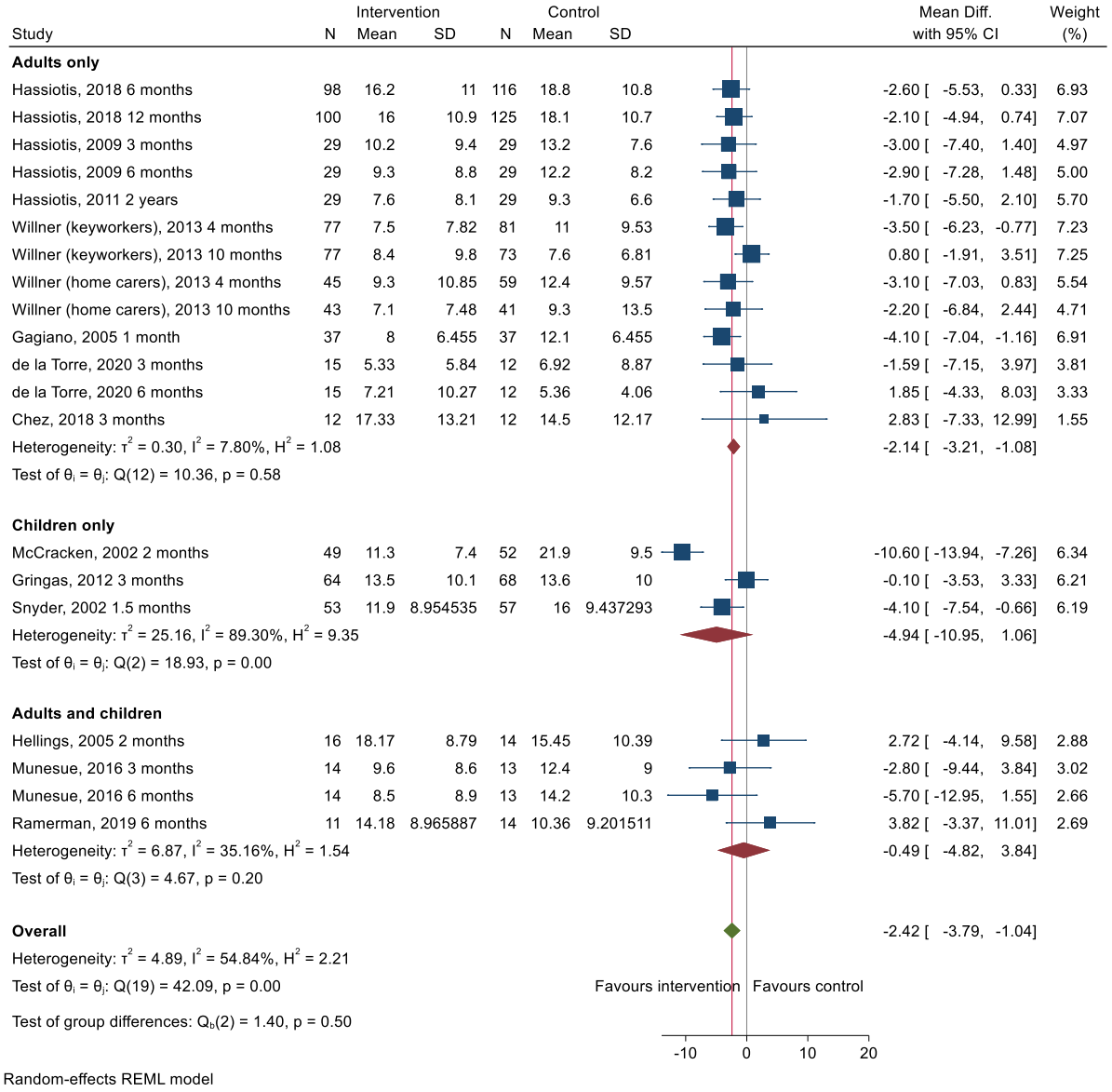
Ramerman et al. (2019)	RCT discontinuation	Children and adults	Pharmacological - risperidone	6 months	N=25	High risk
Chez et al. (2018)	RCT crossover	Adults	Pharmacological - dextromethorphan/quinidine	8 weeks, 3 months	N=14	High risk

1.3.1.1 By study population

There was an advantage for the intervention in studies of adults only (-2.14, 95%CI -3.21, -1.08). However, this subgroup was dominated by three studies (Hassiotis et al., 2009; Hassiotis et al., 2011; Hassiotis et al., 2018; Willner et al., 2013), which included data from more than one time point or rater. There was a wide variation in the estimates for the three studies including children only, with high heterogeneity (I^2=89%). Studies that included adults and children were small and heterogeneous (I^2=35%) (Figure 2). The meta-analysis estimate was not statistically significant with or without Ramerman et al. (2019) (supplementary figure S1).

[figure 2 near here]

Figure 2: Forest plot of ABC-I by study population with Ramerman et al (2019)



1.3.1.2 By follow up time

The ABC-I scores in the intervention arm in the short- and medium-term studies were similar. The mean ABC-I score in the control arm of the short term studies was higher than the mean in the medium term studies. A large difference between intervention and control arms ABC-I scores was shown in studies in children.

The short-term studies show the most variation (I^2=69%) indicating a meta-analysis estimate of -3.05 (95%CI -5.74, -0.36) in favour of the intervention. All longer-term studies were of complex interventions with follow from 10 months to 2 years and with low heterogeneity (I^2=11%). They favoured the intervention, but this was not statistically significant (-1.02, 95%CI -2.77, 0.72) (supplementary figures S2-5).

1.3.1.3 By type of intervention

There is a small advantage for the intervention in the studies of complex interventions -2.09 (95%CI, -3.25, -0.93). The studies of pharmacological interventions were highly heterogeneous (I^2=74%), with a non-significant meta-analysis estimate of -2.71 (95%CI, -5.84, 0.43), influenced by Ramerman et al. (2019). The meta-analysis estimate for interventions without Ramerman et al. (2019) changed to -4.58 (95%CI -8.87, -0.29). In the study that included a combined complex and pharmacological intervention, the meta-analysis estimate was not significant (-0.05, 95%CI -4.19, 4.08) (supplementary figures S6-7).

1.3.1.4 By Risk of Bias

Six studies were rated as “some concerns” for risk of bias and a meta-analysis estimate of -1.96 (95%CI -3.25, -0.66). Seven studies at “high risk” of bias had a meta-analysis estimate of -2.53 (95%CI -5.11, 0.05) with high heterogeneity (I^2=76%) (supplementary figure S8). The meta-analysis estimate without Ramerman et al. (2019) was -3.03 (-5.59, -0.47; supplementary figure S9).

1.3.1.5 Additional considerations

The data in the included studies, indicate a baseline pooled SD of 8.88. Therefore, using a range of effect sizes, the MCID on the ABC-I the difference between interventions achieves values of 4.44 (0.5SD), 2.66 (0.3SD) and 1.78 (0.2SD).

The mean change score for short term studies was -6.25 (SD 4.88), giving a MCID on ABC-I of 2.44. For medium term studies, the change score was -2.45 (SD 1.29), giving an MCID of 0.65.

Sample size calculations based on ABC-I were only available for two studies (Hassiotis et al., 2009; Ramerman et al., 2019). The trial by Hassiotis et al. (2009) was powered to detect an eight point (0.8SD) difference between the intervention and control groups. Ramerman et al.

(2019) set out to detect a 6.5-point change on the ABC-I (equating to approximately 0.77SD with a SD of 8.5).

The ABC-I score change from baseline is shown in table 2.

[Table 2 near here].

Table 2: ABC-I score change by population group, follow up, type of intervention and risk of bias

	Intervention			Control		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
Baseline*	9	17.7	(8.8)	9	18.2	(9.4)
1. Follow up						
Short term*	9	11.7	(8.7)	9	14.0	(9.3)
Medium term**	7	10.3	(9.5)	7	12.0	(8.8)
Long term*	4	9.8	(9.1)	4	11.1	(9.4)
2. Intervention						
Psychosocial**	9	10.2	(9.4)	9	12.4	(9.3)
Pharmacological	9	12.5	(9.0)	9	14.5	(9.6)
Both	2	6.3	(8.1)	2	6.1	(6.5)
*3. Population***						
Adults only	13	10.0	(9.2)	13	11.6	(8.8)
Children only	3	12.2	(8.8)	3	17.2	(9.6)
Adults and children	4	12.6	(8.8)	4	13.1	(9.7)
*4. ROB***						
Some concerns	10	10.3	(9.2)	10	12.4	(8.6)
High	10	11.3	(9.0)	10	13.1	(9.7)

*: number of studies with usable data for baseline, short and long term follow up.

**: It is number of estimates for medium term, intervention type, study population and risk of bias.

1.3.2 Anchor-based

Nineteen psychiatrists, 9 healthcare professionals and 5 family carers rated each item on the ABC-I for all 4 vignettes describing increasing severity of aggressive challenging behaviour

drawn from baseline scores of participants enrolled in a multicentre RCT. All stakeholders were blind to subsequent arm allocation and were asked to rate the change they would expect at 6 months following an intervention (reduction in scores denoting improvement). The mean change score was 12 for vignette 1 (originally scoring 19), and 20 for vignette 2 (originally scoring 29). Other healthcare professionals expected to see the lowest mean score (22) for vignette 3 (originally scoring 39), followed by the psychiatrists at 28 and the family carers at 32. For vignette 4 (highest severity, originally scoring 45), the psychiatrists and healthcare professionals, expected the score to decrease to 30 and 31 respectively, whereas the family carers expected a decrease to 35 (Table 3). Mean percentage difference between baseline and six month scores were approximately 30% for all vignettes and stakeholder groups; notable exceptions to this were vignette 3 where the mean percentage change was a decrease of 43% by healthcare professionals, and 19% by family carers (supplementary table S10).

Table 3: Ratings of change (mean and sd) per vignette and stakeholder group

		Psychiatrists			**Healthcare professionals**			**Family carers**		
Vignette	Baseline score	N	Mean	SD	N	Mean	SD	N	Mean	SD
1	19	14	12.0	(2.4)	6	12.3	(5.4)	5	12.4	(1.5)
2	29	13	19.5	(4.2)	9	20.2	(5.0)	5	20.6	(2.1)
3	39	19	28.3	(5.3)	9	22.4	(8.0)	5	31.6	(5.2)
4	45	19	30.1	(7.9)	8	30.5	(7.2)	4	34.5	(3.0)

1.4 Discussion

We described the estimation of the MCID for a commonly used psychometrically validated scale in studies of aggressive challenging behaviour in people with intellectual disability. We found that distribution and anchor based approaches led to a range of MCID on the ABC-I; notably the former demonstrating smaller changes (range 0.05 to 4.94) compared to the latter (range 6.6 to 16.6). Importantly, pharmacological interventions showed similar magnitude of results to those from complex/psychosocial interventions, but the former covered a range of different medications often non psychotropics but included reduction in aggressive challenging behaviour as a secondary outcome measure.

Considering the severe consequences that such behaviour has for people with intellectual disabilities and their family carers, the views of stakeholders are particularly important. The non-overlapping nature of the findings may be explained by the fact that each vignette described an individual case in isolation whereas in research studies, especially randomised controlled trials, known and unknown confounders are more likely to be balanced. Further, clinical equipoise, which is important in trials, is usually absent in clinical care. However, it is also the case that all the studies that were included in the meta-analysis showed only modest clinical benefits well below what is expected by practitioners and experts by experience. So far, very little has been published about the size of the MCID which widely used interventions for aggressive challenging behaviour that is realistic or important to account for in the design of clinical trials. Bowring et al (2018), previously examined the clinically significant and reliable change scores of the Behavior Problems Inventory – Short Form (BPI-S) with the view of assisting professionals in establishing what change is necessary following an intervention. The authors used the framework by Jacobson and Truax (1991) to create norms for the BPI-S that clinicians could use to demonstrate that a service user had made significant clinical gains by moving from the dysfunctional into the functional domains. However, there are no available data of the BPI-S having been used in clinical trials and the framework used is derived from normative observational data.

As more clinical trials are being carried out in the field of intellectual disability, signal of effect of interventions is an essential part of the advancement of research in this population group. In this effort, introducing the patient-centred concept of MCID is an important contribution to better understand what constitutes meaningful subjective outcomes (McGlothlin & Lewis, 2014).

We believe that the present work underlines the importance of the concept of MCID on an outcome measure and the implications of its use in research and practice. Short of a clear

recommendation, researchers intending to use the ABC-I as primary outcome, should consider as the basis of calculating the sample size on a MCID of 2.5 which is close to the centre of the distribution-based range and the meta-analysis estimate for medium term (approximately 6 months) outcomes at 2.57 (Supplementary Figure S3). RCTs are expensive and of long duration and must adhere to strict ethical and scientific principles. Under-estimating the required sample size and therefore, reporting on the target different apriori including the assumptions behind it are essential reporting requirements. However, it is possible that large enough samples may show statistical significance of a given difference which, though, may be clinically negligible. In the light of the non-overlapping range of the MCID between approaches, researchers and clinicians may wish to place higher value on the perspectives of affected stakeholders and justify that choice.

Pre-existing convention to utilise standardised mean differences may be limited by participant inclusion criteria, between populations and may not be generalizable. It has also been shown that patients are unlikely to endorse improvement where symptoms may have been mild, linking baseline severity to reports of clinical benefit (Kunali et al, 2020). In recognition of the importance of accurate sample size calculations in the conduct of RCTs, Cook et al (2018) produced the DELTA^2 Guidance on how to choose an appropriate MCID and to report on the subsequent sample size calculation. They have developed 10 recommendations for researchers to consider when planning a clinical trial and suggest that the chosen MCID should be important to at least one stakeholder group if seen as likely to achieve or to alter practice.

1.4.1 Strengths and Limitations

This is the first study to investigate the MCID on the widely used ABC-I outcome measure. As most of the included studies did not use the ABC-I as the primary outcome, it was not possible to determine what difference between intervention and control group would have been expected. All the studies were rated as some concern or high risk of bias and larger studies generally showed smaller effect sizes. Risk of Bias categorisations showed that dropout of up to 30%, meant that six small studies had 30 or less participants providing data at the first or only follow-up (Carminati et al., 2016; Chez et al., 2020; de la Torre et al., 2020; Hellings et al., 2005; Munesue et al., 2016; Ramerman et al., 2019). Distribution based approaches for MCID derivation are based on mean differences that do not take into account individual circumstances especially for such a remitting/relapsing condition as aggressive challenging behaviour. Finally, anchor based approaches may be subject to recall bias by patients, and

clinicians or other experts may not be a valid and/or reliable avenue to establishing the MCID on a given measure.

1.4.2 Conclusions

Greater use of the MCID to power pragmatic (effectiveness) trials is important in the interpretation of benefits and harms of a wide range of interventions tailored to decrease challenging behaviour. We believe that this is a fertile area of future investigation where different thresholds are considered for different population subgroups, for example, where there exists higher severity of symptoms at baseline and multi/comorbidity. Added clarity about the impact of an intervention can also be provided by considerations of costs and adverse events which are all very important aspects of research and clinical practice and relevant to both patient perspective and evidence based care. Finally, more research is needed on MCIDs for other outcome measures in the field of intellectual disabilities so that the differences in range found between methods in the current study can be understood more fully.

Word count: 6311

Declaration of interest

None

Funding

The PETAL programme is funded by the National Institute for Health Research (NIHR) [PGfAR NIHR200120]. The views expressed within the manuscript are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Acknowledgments

We thank Maria Lahab, Research Assistant, who and carried out the literature search study quality assessment and developed initial draft of the manuscript.

Author contribution

AH formulated the research question, wrote and substantially revised the manuscript and is the guarantor of the project; LM carried out the meta-analysis and contributed to the systematic review and manuscript content; RH, SAC, AJ, CM, RH, AS, PR, ES, AA, VC discussed interpretation of findings and contributed to iterations of the manuscript.

Data availability

The data that support the findings of this study are available from the corresponding author (AH) upon reasonable request.

References

Aman, M.G., De Smedt, G., Derivan, A., Lyons, B., & Findling, R.L. (2002) Risperidone Aggressive Behavior Study Group. Double-blind, placebo-controlled study of risperidone for the treatment of aggressive behaviors in children with subaverage intelligence. *American Journal of Psychiatry,159*, 1337-1346.

Aman, M.G., & Singh, N.N. (2017) *Aberrant Behavior Checklist Manual* (2nd ed.). East Aurora, NY: Slosson Educational Publications.

Arora, S., Goodall, S., Viney, R., & Einfeld, S. (2020). Societal cost of childhood intellectual disability in Australia. *Journal of Intellectual Disability Research*, *64*, 524-537.

Bowring DL, Totsika V, Hastings RP, Toogood S (2018). Toward data-based clinical decision making for adults with challenging behavior using the Behavior Problems Inventory-Short Form (BPI-S), *Tizard Learning Disability Review*, 23, 103-110

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, … Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory--II according to the patient's perspective. *Psychological Medicine*, *45*, 3269–3279.

Carminati, G. G., Gerber, F., Darbellay, B., Kosel, M. M., Deriaz, N., Chabert, J., … Carminati, F. (2016). Using venlafaxine to treat behavioral disorders in patients with autism spectrum disorder. *Progress in Neuro-psychopharmacology & Biological Psychiatry*, *65*, 85–95.

Chez, M., Kile, S., Lepage, C., Parise, C., Benabides, B., & Hankins, A. (2020). A Randomized, Placebo-Controlled, Blinded, Crossover, Pilot Study of the Effects of Dextromethorphan/Quinidine for the Treatment of Neurobehavioral Symptoms in Adults with Autism. *Journal of Autism and Developmental Disorders*, *50*, 1532–1538.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed*). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers

Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA et al (2018). DELTA^2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ* 2018;363:k3750

Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., Jr, & Schuler, T. C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *The spine journal: official journal of the North American Spine Society*, *7*, 541–546.

de la Torre, R., de Sola, S., Farré, M., Xicota, L., Cuenca-Royo, A., Rodriguez, J., … TESXF Study Group (2020). A phase 1, randomized double-blind, placebo controlled trial to evaluate safety and efficacy of epigallocatechin-3-gallate and cognitive training in adults with Fragile X syndrome. *Clinical Nutrition, 39*, 378–387.

Forkmann, T., Scherer, A., Boecker, M., Pawelzik, M., Jostes, R., & Gauggel, S. (2011). The Clinical Global Impression Scale and the influence of patient or staff perspective on outcome. *BMC Psychiatry,11*, 83.

Fujiura, G. T., Li, H., & Magaña, S. (2018). Health services use and costs for Americans with intellectual and developmental disabilities: a national analysis. *Intellectual and Developmental Disabilities*, *56*, 101-118.

Gagiano, C., Read, S., Thorpe, L., Eerdekens, M., & Van Hove, I. (2005). Short- and long-term efficacy and safety of risperidone in adults with aggressive behavior disorders. *Psychopharmacology*, *179*, 629–636.

Glover, G., Williams, R., Heslop, P., Oyinlola, J., & Grey, J. (2017). Mortality in people with intellectual disabilities in England. *JIDR*, *61*, 62-74.

Griffith GM, Hastings RP (2014). 'He's hard work, but he's worth it'. The Experience of Caregivers of Individuals with Intellectual Disabilities and Challenging Behaviour: A Meta-Synthesis of Qualitative Research. *JARID*, 27, 401-19

Gringras, P., Gamble, C., Jones, A. P., Wiggs, L., Williamson, P. R., Sutcliffe, A., Appleton, R., on behalf of the MENDS Study Group. (2012). Melatonin for sleep problems in children with neurodevelopmental disorders: randomised double masked placebo controlled trial. *BMJ 345*, e6664.

Hassiotis, A., Robotham, D., Canagasabey, A., Romeo, R., Langridge, D., Blizard, R., … King, M. (2009). Randomized, single-blind, controlled trial of a specialist behavior therapy team for challenging behavior in adults with intellectual disabilities. *American Journal of Psychiatry*, *166*, 1278–1285.

Hassiotis, A., Canagasabey, A., Robotham, D., Marston, L., Romeo, R., & King, M. (2011). Applied behaviour analysis and standard treatment in intellectual disability: 2-year outcomes. *The British Journal of Psychiatry, 198*, 490–491.

Hassiotis, A., Poppe, M., Strydom, A., Vickerstaff, V., Hall, I. S., Crabtree, J., … Crawford, M. J. (2018). Clinical outcomes of staff training in positive behaviour support to reduce challenging behaviour in adults with intellectual disability: cluster randomised controlled trial. *The British Journal of Psychiatry, 212*, 161–168.

Hays, R. D., & Woolley, J. M. (2000). The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it?. *PharmacoEconomics*, *18*, 419–423.

Hellings, J. A., Weckbaugh, M., Nickel, E. J., Cain, S. E., Zarcone, J. R., Reese, … Cook, E. H. (2005). A double-blind, placebo-controlled study of valproate for aggression in youth with pervasive developmental disorders. *Journal of Child and Adolescent Psychopharmacology*, *15*, 682–692.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (editors). (2020). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Retrieved from https://training.cochrane.org/handbook accessed January 2021

Howard, R., Phillips, P., Johnson, T., O'Brien, J., Sheehan, B., Lindesay, J., … Gray, R. (2011). Determining the minimum clinically important differences for outcomes in the DOMINO trial. *International Journal of Geriatric Psychiatry*, *26*, 812–817.

Hughes-McCormack, L., Greenlaw, N., McSkimming, P., McCowan, C., Ross, K., McConnachie, A., et al. (2021) Trends over time in the management of long-term conditions in primary health care for adults with intellectual disabilities, and the health care inequality gap

with the general population. *Journal of Applied Research in Intellectual Disabilities*, 34, 634-47.

Jacobson NS, Follette, WC, Revenstorf D (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336–352

Jacobson NS, Truax P (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19

Kline Leidy N, Wyrwich W K (2005) Bridging the Gap: Using Triangulation Methodology to Estimate Minimal Clinically Important Differences (MCIDs), *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 2:1, 157-165

Lunsky, Y., De Oliveira, C., Wilton, A., & Wodchis, W. (2019). High health care costs among adults with intellectual and developmental disabilities: a population-based study. *Journal of Intellectual Disability Research*, *63*, 124-137.

McGlothlin AE, Lewis RJ (2014). Minimal clinically important difference: defining what really matters to patients. *JAMA* 312, 1342-3.

McCracken, J. T., McGough, J., Shah, B., Cronin, P., Hong, D., Aman, M. G., … Research Units on Pediatric Psychopharmacology Autism Network. (2002). Risperidone in children with autism and serious behavioral problems. *The New England Journal of Medicine*, *347*, 314–321.

McGuire, D. O., Tian, L. H., Yeargin-Allsopp, M., Dowling, N. F., & Christensen, D. L. (2019). Prevalence of cerebral palsy, intellectual disability, hearing loss, and blindness, National Health Interview Survey, 2009–2016. *Disability and Health Journal*, *12*, 443-451.

Munesue, T., Nakamura, H., Kikuchi, M., Miura, Y., Takeuchi, N., Anme, T., … Higashida, H. (2016). Oxytocin for Male Subjects with Autism Spectrum Disorder and Comorbid Intellectual Disabilities: A Randomized Pilot Study. *Frontiers in Psychiatry*, *7*, 2.

National Institute for Health Care and Excellence. (2015). *Challenging Behaviour and Learning Disabilities: Prevention and Interventions for Children with Learning Disabilities Whose Behaviour Challenges* (NICE Guideline 11 Methods). [PDF]. Retrieved from https://www.nice.org.uk/guidance/ng11 accessed May 2021

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care*, *41*, 582–592.

Ramerman, L., de Kuijper, G., Scheers, T., Vink, M., Vrijmoeth, P., & Hoekstra, P. J. (2019). Is risperidone effective in reducing challenging behaviours in individuals with intellectual disabilities after 1 year or longer use? A placebo-controlled, randomised, double-blind discontinuation study. *Journal of Intellectual Disability Research*, *63*, 418–428.

Sedaghat, A. R. (2019). Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. *Otolaryngology--head and neck surgery: official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *161*, 551–560.

Snyder, R., Turgay, A., Aman, M., Binder, C., Fisman, S., & Carroll, A. (2002). Effects of risperidone on conduct and aggressive behavior disorders in children with subaverage IQs. *Journal of the American Academy of Child Adolescent Psychiatry*, *41*,1026-1036.

Sterne, J., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., … Higgins, J. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed.)*, *366*, l4898.

Willner, P., Rose, J., Jahoda, A., Stenfert Kroese, B., Felce, D., MacMahon, P., … Hood, K. (2013). A cluster randomised controlled trial of a manualised cognitive behavioural anger management intervention delivered by supervised lay therapists to people with intellectual disabilities. *Health Technology Assessment 17*(21), 1–vi.

Wright, A., Hannon, J., Hegedus, E. J., & Kavchak, A. E. (2012). Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *The Journal of Manual & Manipulative Therapy*, *20*, 160–166.

Zablotsky, B., Black, L. I., Maenner, M. J., Schieve, L. A., Danielson, M. L., Bitsko, R. H., ... & Boyle, C. A. (2019). Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. *Pediatrics*, *144*(4), e20190811.