A systematic review of fidelity measurements in complex interventions for people with intellectual disabilities and behaviours that challenge

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Advances in Mental Health and Intellectual Disabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>AMHID-09-2018-0040.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Research Paper</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Fidelity, Complex intervention, Intellectual disability, Behaviours that challenge, Positive Behaviour Support, Randomised controlled trial</td>
</tr>
</tbody>
</table>
Introduction

There are approximately 26,500 adults with intellectual disabilities (ID) and behaviours that challenge in England (Emerson et al, 2011). Behaviours that challenge are enduring and often have serious negative consequences for the person (e.g. administration of medication, abuse or in-patient admissions) and for those who support them (e.g. physical/mental health problems or high job turnover) (Emerson and Enfield, 2011; Allen, 2009; Sturmey, 2009).

Currently, Positive Behaviour Support (PBS) is considered the best practice for supporting individuals with ID who engage in behaviours that challenge (Department of Health, 2014; British Psychological Society, 2018; National Institute for Clinical Excellence, 2015). PBS is a multicomponent biopsychosocial intervention that aims to improve individuals’ quality of life and eventually reduce the display of behaviours that challenge (PBS competence framework, 2015). PBS evolved from Applied Behaviour Analysis (ABA) that forms a central core of PBS and provides the framework and techniques for understanding and reducing behaviours that challenge (Allen, 2009). Behaviour support plans are developed as part of PBS to provide a step by step guide to carers of how to support an individual with ID effectively (Gore et al, 2015; Allen et al, 2005; LaVigna and Willis, 2005).

There is evidence from systematic and meta-analytic reviews of studies using N=1 designs that interventions based on PBS can produce reductions in behaviours that challenge up to 50% (Carr et al, 1999; Dunlap et al, 2008; LaVigna and Willis, 2012). One UK-based RCT showed that behaviours that challenge had reduced by 43% after the implementation of PBS (Hassiotis et al, 2009). However, only a small proportion of those who display behaviours that challenge receive behavioural support in England (Emerson et al, 2000; Allen et al, 2005). Possible reasons as to why behavioural support is not widely available are: there is a lack of services available in the UK based on PBS principles, small numbers of staff...
members are trained in PBS and there are long waiting lists in the services that provides PBS/ABA (Toogood, 2015; Allen et al, 2005; Beadle et al, 2006). Davidson et al (2015) found that the current provision of peripatetic support for individuals with behaviours that challenge in the UK is very low; only 46 services identified themselves as such and out of these only 47% had reported providing support for behaviours that challenge based on PBS principles. These findings are in contract to the main focus of PBS which aims to apply research-validated principles in real-life settings where individuals with ID live (Carr et al, 2002).

Fidelity data in clinical trials of PBS could provide useful insights how to improve the application of PBS in everyday settings (Boutron et al, 2008; Craig et al, 2008). Fidelity, which is also known as integrity or adherence, refers to the extent an intervention is implemented and delivered as indicated by the intervention developers (Moore et al, 2015; Craig et al, 2008). Fidelity measurements are particularly important for complex interventions because there are multiple factors that affect its delivery and efficacy (Craig et al, 2008). According to the Medical Research Council (MRC) guidelines, complex interventions are defined as having multiple active components that influence outcomes, multiple agents with a range of different skills are involved in the delivery and the intervention is delivered in real life contexts (Craig et al, 2008).

PBS is a complex intervention and as such has multiple interacting elements that are important for the reduction of behaviours that challenge, it is constantly adapted to take into account the service users’ strengths and targets the person’s environment (Carr et al, 2002). Therefore, it is plausible that during the implementation of PBS several elements of this intervention might be omitted or added (Carr et al, 2002). Besides, multiple other variables (such as therapists’ competence, skills, motivation, time commitments, goodness of fit,
management support, staff-turnover rates) also impact its delivery and implementation 
(Hassiotis et al, 2018). Examining fidelity data reported in PBS trials are not only necessary 
for the internal and external validity of the study but it also provides information about 
esential elements of PBS that need to be implemented as identified in the protocol and of 
those that can be adapted without decreasing its effectiveness (Carroll et al, 2007); it 
highlights the aspects of PBS that service providers find difficult to deliver, pointing to 
modifications in the administration of PBS and what additional training is needed in order to 
improve its implementation (Spee et al, 2014; Mowbray et al, 2003).

Systematic reviews report that fidelity measurements in the behavioural analysis literature are 
examined fidelity measurements reported in school-based experimental (within-group or 
single-case design) studies with children published in the Journal of Applied Behavior 
Analysis between 1991 to 2005. Researchers found that only 39% of studies provided any 
mention of fidelity assessments; 30% reported quantitative data (in the form of percentage of 
implementation) and 9% of studies mentioned that fidelity measurements were completed but 
quantitative data were not provided in the papers (McIntyre et al, 2007). Another review 
(Sanetti et al, 2012) found similar findings: 42% of studies with children published in the 
Journal of Positive Behavior Interventions between 1999 and 2009 included quantitative 
fidelity data and 7% of studies mentioned fidelity assessments without providing quantitative 
data (McIntyre et al, 2007; Sanetti et al, 2012). These reviews indicate that fidelity 
measurements in behavioural literature are not new but the low rates of reporting fidelity 
measurements in the studies over the years remain stable. Nevertheless, these systematic 
reviews examined fidelity in single-case design studies published in only one journal and 
often excluding people with ID from their samples. Different trends of fidelity measurements
might be found in the literature using different inclusion criteria, considering more than one
medical journal and including people with ID.

With this in mind, the primary aim of this study was to carry out a systematic review
examining the approaches used to measure fidelity in randomised controlled trials of complex
interventions for behaviours that challenge which are based on PBS/ABA principles.

**Methods**

**Search strategy**

Searchers of the electronic databases MEDLINE, Embase, PsycINFO, Web of Science and
CINAHL Plus were conducted from April 2017 to June 2017. The first 50 pages of grey
literature were also searched and sources included: Research gate, Google scholar, Google,
NICE Evidence Search, The King’s Fund, Zetoc, Proquest, WorldCat, OpenGrey and Clinical
Trials.gov. We also hand searched the reference lists of the studies included in the review.
The search terms related to intellectual disability, behaviours that challenge, fidelity and
randomised controlled trial were combined using the Boolean operator “AND” (see
supplementary material 1 for the full search strategy).

**Inclusion criteria**

*Participants:* children or adults with mild, moderate, severe or profound ID and behaviours
that challenge. ID had to be explicitly identified in the studies such as an IQ below 70
alongside a standardised measure or diagnostic criteria; behaviours that challenge also had to
be measured by standardised questionnaires.
Studies, which had participants with other neurodevelopmental comorbidities in addition to ID and behaviours that challenge (e.g. Autistic Spectrum Disorders), were included only if at least 50% of the sample had ID.

**Intervention:** complex interventions for behaviours that challenge which are based on PBS or ABA approaches (e.g. pivotal response training or differential reinforcement). Interventions that did not target behaviours that challenge were excluded.

**Comparison:** any control group (e.g. treatment as usual, no treatment, waitlist control or any alternative management strategy).

**Outcomes:** studies had to include a statement about measuring and reporting fidelity. The approach proposed by Dane and Schneider (1998) was used to appraise the dimensions of fidelity reported in the studies (see Table 1 for definitions).

**Study design:** randomised controlled trials.

**Publication:** published articles were limited to the English language covering the period from January 1990 to Jun 2017.

**Table 1: Fidelity components (based on Dane and Schneider, 1998)**
Study selection

Titles and abstracts were screened by one reviewer (LP) against the inclusion criteria. Full text articles of potentially eligible studies were obtained and assessed by two reviewers (LP and AH). Numbers of excluded studies and its reasons were recorded. Any disagreements or discrepancies about included studies were resolved by a consensus and discussion with a third reviewer (AA). Two reviewers (LP and AH) then independently extracted data into a standardised form that was developed for this study.

Risk of bias in included studies

The Cochrane risk of bias tool (Higgins and Green, 2011) was used independently by two reviewers (LP and AA) in order to detect risk of bias in each study. Reviewers rated risk of bias as low (+), high (-) or unclear (?) for each of the seven domains: allocation, selection, performance, detection, attrition, reporting and other biases (see Table 2 for definitions).

Table 2: Risk of bias (Higgins and Green, 2011)
Results

Electronic and hand searches identified 3482 records. Figure 1 shows a flow diagram of the study selection process for this review. After removal of duplicates, 3439 unique articles were screened and 36 full-text papers of potentially relevant articles were retrieved. Thirty-one articles did not meet inclusion criteria (see supplementary material 2 for a list of excluded articles). Thus, five articles were included in the review. Table 3 shows an overview of the studies.

Figure 1: A PRISMA flow chart of study selection process.
Table 3: Characteristics of the included studies
Table 4: Elements of fidelity reported in the included studies
Overview of the included studies

Study design and location

The current review identified five small sample studies of which one was randomised controlled trial (Singh et al, 2016) and four were pilot clinical trials (Ingersoll et al, 2016; Johnson et al, 2013; Hassiotis et al, 2009; Reitzel et al, 2013). Studies were published between 2009 and 2016 in the UK, USA and Canada. Participants were recruited from the following settings: residential homes, inpatient, community intellectual disabilities service and education. An overview of studies is provided in Table 3.

Participants

The total number of participants of the five included studies was 178. The number of participants in the studies ranged from 15 to 45. Majority of participants in the included studies were males apart from one study which did not specify numbers of males and females in their sample (Reitzel et al, 2013). Two studies focused on adults (Singh et al, 2016; Hassiotis et al, 2009), the other two on children (Johnson et al, 2013; Reitzel et al, 2013) and one on adolescents (Ingersoll et al, 2016). Participants in the included studies had various levels of ID (mild to severe) and behaviours that challenge. Three studies had participants with Autistic Spectrum Disorders or Pervasive Developmental Disorders, Not Otherwise Specified in addition to ID (Ingersoll et al, 2016; Johnson et al, 2013; Reitzel et al, 2013).

Interventions

All interventions were for behaviours that challenge based on ABA or PBS principles and had some overlap in terms of the content of techniques used (e.g. functional assessments, antecedent manipulation, data-collection, extinction, reinforcement and skills teaching).

One study examined individual reciprocal imitation training in addition to treatment as usual against treatment as usual which consisted of PBS intervention and individual education sessions (Ingersoll et al, 2016). Singh et al (2016) compared mindfulness-based PBS training
to a control group which received PBS training alone. Johnson et al (2013) examined
behavioural parent training in addition to psychoeducational programme against treatment as
usual (which consisted of psychoeducational programme). Hassiotis et al (2009) assessed
behaviour therapy (applied behaviour analysis and positive behavioural support) against
standard care alone. Reitzel et al (2013) study compared functional behaviour skills training
to a waitlist-control group.

Outcomes

See Table 4 for the characteristics of fidelity elements reported in the studies.

Adherence of implementation

Steps taken to measure adherence were inconsistently reported in the studies. Only two
studies provided detailed evaluations of whether an intervention was delivered as described in
the protocol which included: records of methodological techniques used, interrater-reliability
for the adherence measures and quantitative adherence data (Ingersoll et al, 2016; Johnson et
al, 2013).

In Ingersoll’s et al (2016) study six randomly selected intervention sessions were assessed by
researchers using fidelity checklists. The agreement between researchers for fidelity
checklists was 0.96 which was calculated on 33% of the sessions using intraclass correlation
coefficients. Average fidelity rating for the intervention was 91.21% (range 90.37–100%)
indicating that teachers implemented the reciprocal imitation intervention, on average, as
described by the protocol. However, there was no information provided in the paper what
these fidelity checklists were capturing.

Johnson’s et al (2013) study used both fidelity checklists and 10% of randomly selected
videotaped sessions to assess the therapists’ adherence. This study was the only one that
provided fidelity definition that authors are using in the paper. The fidelity checklist included therapists’ adherence to the protocol (whether all essential elements were provided) and parents’ adherence to the intervention (whether parents completed homework or implemented suggested strategies). Therapists were asked to use fidelity checklists after each session and rate themselves on a scale 0 to 2 whether they achieved the goals of the intervention, indicating 0 if the goals were not achieved, 1 if goals were partially achieved or 2 if goals were achieved. Additionally, all intervention sessions were videotaped and 10% of the sessions were randomly selected and assessed by independent observers. Average Interrater-reliability for the randomly selected sessions was 98.2% (range 81.8–100%) and for parent adherence was 91.8% (range 66.7–100%). Mean fidelity rating for the Behavioural Parent Intervention was 98% (range 83–100%) and for parent adherence 93% (range 75–100%).

Partially reported adherence was found in Hassiotis’s et al (2009) and Reitzel’s et al (2013) studies. Hassiotis et al (2009) indicated that fidelity for implementation of the intervention was high without mentioning methodological techniques used and quantitative data. Reitzel et al (2013) study only reported that staff completed fidelity checklists without providing quantitative data in the paper.

**Dose**

The amount of an intervention delivered was reported in detail as all five studies provided details of the frequency and duration of an intervention (Ingersoll et al, 2016; Johnson et al, 2013; Hassiotis et al, 2009; Reitzel et al, 2013; Singh et al, 2016). However, only one study provided the number of sessions that participants had actually completed (Johnson et al, 2013). In this study, self-reports revealed that parents’ attendance to the intervention sessions was 97.3%.
**Quality of delivery**

The most frequently reported measure of assurance of the quality of delivery was training in the intervention techniques and choice of training providers based on qualifications and experience (all five studies). However, only Reitzel et al (2013) study examined parents’ knowledge of ABA principles and sense of competence before and after the training. None of the other included studies examined whether staff or family members acquired necessary knowledge and skills after the training.

Ingersoll et al (2016) and Johnson et al (2013) had also mentioned that supervision was available for intervention providers. Training fidelity checklists (whether the training was provided as indicated in the protocol) only mentioned in Ingersoll’s et al (2016) and Singh’s et al (2016) studies. Even though Singh et al (2016) videotaped 10 random parts of the training and reported quantitative training fidelity data (which was indicated as 100%), only one person was responsible for rating the quality of the training and it was not reported whether this assessor was independent of the study team or not. Thus, it is unclear whether subjective opinion biased the results.

**Responsiveness**

Measurement of participants’ engagement into the content of the intervention and their perceptions about the usefulness of an intervention were rarely reported. Two studies out of five reported participants’ responsiveness. Ingersoll et al (2016) study explored intervention providers’ views about the benefits of the intervention and whether participants enjoyed the sessions. However, participants’ views about the usefulness of the intervention have not been explored directly in this study. In Johnson’s et al (2013) study parents received satisfaction questionnaires to rate how useful they found teaching goals and specific intervention strategies.
Program differentiation

None of the included studies carried out a component analysis or qualitative research to examine which elements of an intervention are essential and the most effective.

Risk of bias in the included studies

See Figure 2 (below) for a summary of risk of bias in the included studies and also supplementary material 3 for reviewers’ judgments of risk of bias. Overall, the included studies were found to be at high risk of reporting bias especially for these domains: allocation concealment, blinding, incomplete outcome data, selective reporting and other bias.

Four out of five studies had reported specific mechanisms used to produce random sequence generation (Ingersoll et al, 2016; Singh et al, 2016; Hassiotis et al, 2009; Reitzel et al, 2013).

Two studies had provided information about allocation concealment (Singh et al, 2016; Johnson et al, 2013). Blinding of participants, personnel and outcome assessors were stated in one study (Reitzel et al, 2013). Missing outcome data and its reasons were reported in four studies (Ingersoll et al, 2016; Singh et al, 2016; Hassiotis et al, 2009; Reitzel et al, 2013). Selective reporting was evident in two studies which indicated that fidelity measurements were completed but quantitative data were not provided (Hassiotis et al, 2009; Reitzel et al, 2013). All five studies had other potential threats to validity which were: small sample sizes reported in all studies (range 15 to 45 participants), longer sessions in the treatment condition compared to the control condition in one study (Ingersoll et al, 2016), administration of reciprocal imitation training in addition to PBS plans which also target communication as a part of a plan (Ingersoll et al, 2016) and participants’ baseline characteristics unbalanced between the groups (Reitzel et al, 2013).
Discussion

The review presents data from five RCTs that had reported measurements of fidelity in complex interventions for behaviours that challenge based on PBS/ABA principles. We found variable and inconsistent descriptions of how fidelity was measured and reported in the studies. The most frequently provided fidelity elements found in four out of five studies were adherence of implementation, dose and some aspects of quality of delivery (Ingersoll et al, 2016; Johnson et al, 2013; Hassiotis et al, 2009; Reitzel et al, 2013). However, there were considerable differences between studies about how these fidelity elements were analysed and reported.
Only two studies in the review had provided detailed accounts on how adherence of implementation was measured reporting methodological techniques, interrater reliability and quantitative fidelity data (Ingersoll et al., 2016; Johnson et al., 2013). By contrast, the other two studies only mentioned that adherence of implementation was high without providing measurement techniques used and quantitative data (Hassiotis et al., 2009; Reitzel et al., 2013). Even though frequency and duration of the intervention sessions were reported quite well, only one study indicated the number of sessions that participants had actually completed (Johnson et al., 2013). Only Reitzel et al (2013) examined whether staff and family members acquired skills and knowledge after the training. One study in the review only mentioned steps taken to measure training fidelity (whether training was provided as indicated in the protocol) and did not report any other fidelity elements (Singh et al., 2016). None of the studies assessed program differentiation to examine which components of the intervention are essential and the most effective. Only one study reported the definition of fidelity that authors are using in the paper (Johnson et al., 2013).

Besides inconsistent fidelity measurements found in the review, there is room for improvement for studies which had reported some fidelity measurements. Ingersoll et al (2016) study only used one methodological technique to measure adherence of implementation and it is accepted that several techniques need to be applied to provide reliable data (Spilane et al, 2007). Even though Johnson et al (2013) used several methods to collect quantitative data, intervention providers rated themselves using fidelity checklists whether they had implemented the intervention as indicated in the protocol. Therapists may be afraid to disclose implementation difficulties and may not respond truthfully leading to inaccurate fidelity data. Moreover, Singh et al (2016) reported that training fidelity was 100% suggesting that training was provided exactly as indicated in the protocol. However, only one
observer was involved in the measurement of training fidelity and these findings should be interpreted with caution.

We found a paucity of research measuring fidelity in complex interventions for behaviours that challenge based on PBS/ABA principles and this finding is in line with previous literature (McIntyre et al, 2007; Sanetti et al, 2012). Our review included five papers that met the inclusion criteria and this number is lower than McIntyre’s et al (2007) and Sanetti’s et al (2012) studies found (46 and 35 studies respectively). However, these reviews used different inclusion criteria and involved studies with designs other than RCTs. The limited number of RCTs found in the review reflects a general lack of RCTs including people with intellectual disabilities (Beavis et al, 2007; Hassiotis, 2009).

Reasons for poor fidelity measurements in the behavioural literature are unclear. One possible reason could be that there are multiple definitions of fidelity used in the literature and no consensus on which fidelity elements have to be assessed in complex interventions for behaviours that challenge (Bellg et al, 2004). In the review, we used the most commonly applied fidelity framework across different disciplines (Dane and Schneider, 1998), to appraise fidelity elements (Carroll et al, 2007; Century et al, 2010). However, it is unclear whether all these five fidelity elements indicated in the framework are equally important and need to be evaluated in complex interventions for behaviours that challenge in order to provide a comprehensive fidelity measurement. The CONSORT and MRC guidelines on reporting findings from clinical trials indicate the importance of measuring fidelity in clinical trials but there is no recommendation how to do so (Boutron et al, 2008; Craig et al, 2008). A standardised fidelity checklist that could be easily applied in different settings does not exist and usually such scales are developed for each study (Spilane et al, 2007).
Strengths and limitations

The strengths of this study include a systematic review of fidelity measurements in complex interventions for behaviours that challenge in people with ID that has not been previously conducted in this area. We used a comprehensive search strategy and reviewed grey literature such as conference reports and clinical trial registration websites to find unpublished studies.

Nevertheless, the findings of this study should be interpreted with caution. Firstly, the review included studies published only in the English language and there is a risk that we did not capture all relevant studies. Secondly, we reviewed fidelity measurements in a variety of different behavioural interventions for behaviours that challenge which are mainly based on PBS or ABA principles and this mixture of studies reflects the lack of RCTs carried out including people with behaviours that challenge. Thus, we might focused on the fidelity elements that may not be equally important to look at ABA and PBS interventions. Thirdly, we could not do a meta-analysis because the fidelity measurements reported in the studies were inconsistent and only two studies had provided quantitative data (Ingersoll et al, 2016; Johnson et al, 2013).

Implications for research and practice

Our review confirmed the limited number of RCTs examining complex interventions for behaviours that challenge based on PBS/ABA principles despite it being an important criterion of the study’s internal and external validity. Fidelity is being seen as the degree to which the intervention was implemented as indicated in the protocol (Moore et al, 2015). However, for complex interventions nonspecific therapy factors such as therapist’s competence, skills, motivation, communication style may play an important part how an intervention is implement and might need to be considered as well (Chatoor and Kurpnick, 2001).
We therefore suggest that future studies should clarify which fidelity elements have to be assessed in complex interventions for behaviours that challenge in people with ID. This information might ease the current confusion in the literature and improve the measurements of fidelity in the clinical trials. The development of a standardised approach for measuring fidelity in such interventions for behaviours that challenge could be the next step aiming to increase fidelity measurements in the trials. Our findings from the review could be used as a starting point for researchers to look at methodological techniques reported in complex interventions in this population and to see if any of the helpful techniques could be used. Besides, the Consort and MRC guidelines could establish the best practice for how to measure fidelity in complex interventions which might provide clarity for many researchers how to measure fidelity in PBS/ABA trials.

PBS is a complex intervention and as such, it is imperative that it is subject to the same quality assurance as other interventions. Given that its delivery can be variable, meaning that therapists may implement deviations that are likely to compromise its delivery, it is necessary to develop a standardised approach for measuring fidelity in research (Denne et al, 2013). PBS is about applying research-validated interventions into everyday settings (Carr et al, 2002). However, without fidelity data widespread PBS application into real-world contexts is challenging.
References

References of studies included in the review


Additional references


LaVigna, GW and Willis, TJ (2005), ‘A positive behavioural support model for breaking the barriers to social and community inclusion’, *Tizard Learning Disability Review*, vol. 10, no. 2, pp.16-23.

*Journal of Intellectual and Developmental Disability*, vol. 37, no. 3, pp. 221-231.

McIntyre, LL, Gresham, FM, DiGennaro, FD and Reed, DR, (2007), ‘Treatment Integrity of 
School-Based Interventions with Children in the Journal of Applied Behavior Analysis 1991- 

Moore, GF, Audrey, S, Barker, M, Bond, L, Bonell, C, Hardeman, W, Moore, L, Cathain, 

Mowbray, CT, Holter, MC, Teague, GB and Bybee, D, (2003), ‘Fidelity criteria: 
development, measurement, and validation’, *American Journal of Evaluation*, vol. 20, no. 3, 
pp. 315–340.

National Institute for Clinical Excellence (NICE), (2105), *Challenging behaviour and 
learning disabilities: Prevention and interventions for people with learning disabilities whose 
behaviour challenges (NG11)*. London: NICE

competence framework*.

Sanetti, LMH, Dobey, LM and Gritter, KL, (2012), ‘Treatment Integrity of Interventions 
With Children in the Journal of Positive Behavior Interventions From 1999 to 2009’, *Journal 
of Positive Behavior Interventions*, vol. 14, no. 1, pp. 29–46.

Spee, GA, Polo, AJ and Budd, KS, (2014), ‘Establishing Treatment Fidelity in Evidence- 
Based Parent Training Programs for Externalizing Disorders in Children and Adolescents’, 


Table 1: Fidelity components (based on Dane and Schneider, 1998)

<table>
<thead>
<tr>
<th>Fidelity elements</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherence</td>
<td>The extent to which an intervention content, methods and activities are implemented as described in the protocol. For example, if an intervention provider covered 15 out of the 30 content areas of an intervention, the adherence to the intervention would be 50%.</td>
</tr>
<tr>
<td>Dose (also known as exposure)</td>
<td>The amount of an intervention received by participants compared to the amount of the intervention indicated in the protocol. Dose can include number of sessions completed, frequency and duration of sessions. Ideally, dose should be measured using a combination of different measurements such as attendance logs or checklists, self-reports completed by intervention providers and observations of randomly selected sessions in order to assess the proportion of an intervention delivered (Spillane et al, 2007).</td>
</tr>
<tr>
<td>Quality of delivery</td>
<td>The measurement of the effort required to deliver an intervention. Aspect of quality of delivery can include: the assurance of quality of training, materials and support to those delivering an intervention, provider confidence and interaction style.</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>The degree to which participants engage in the activities and content of an intervention. This can include: participants’ interest in the program, perceived benefits of an intervention or engagement in activities.</td>
</tr>
<tr>
<td>Program differentiation</td>
<td>The degree to which unique components of an intervention are differentiated from each other and from other interventions. Component analysis can help to determine which elements of an intervention are essential and which elements of an implemented intervention cannot be incorporated in the control condition.</td>
</tr>
</tbody>
</table>
Table 2: Risk of bias (Higgins and Green, 2011)

<table>
<thead>
<tr>
<th>Risk of bias</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Allocation bias was examined by looking at whether the method used to assign participants to the groups was truly random (for example, a flip of coin or computerised sequences).</td>
<td></td>
</tr>
<tr>
<td>2. Selection bias was examined by looking at whether anyone in the study could predict allocation to the treatment group and attempts involved to conceal it.</td>
<td></td>
</tr>
<tr>
<td>3. Performance bias was examined by looking at whether intervention providers and participants were blinded to treatment allocation.</td>
<td></td>
</tr>
<tr>
<td>4. Detection bias was examined by looking at whether outcome assessors were blinded to treatment allocation.</td>
<td></td>
</tr>
<tr>
<td>5. Attrition bias was examined by looking at proportion of people in both conditions that stopped having treatment and reasons for it (drop-outs and withdrawal).</td>
<td></td>
</tr>
<tr>
<td>6. Reporting bias was examined by looking at whether more outcomes were measured than were reported.</td>
<td></td>
</tr>
<tr>
<td>7. Other potential threats to validity not addressed by these six domains were stated in this section.</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Characteristics of the included studies

<table>
<thead>
<tr>
<th>Study, location and design</th>
<th>Sample</th>
<th>ID level and its measure</th>
<th>CB and its measure</th>
<th>Intervention</th>
<th>Comparison</th>
</tr>
</thead>
</table>
| **Hassiotis et al, 2009; UK, Pilot RCT** | Community ID service; 63 adults with severe ID and CB.  
*Intervention group*: 32 participants (19 males and 13 females) Mean age = 39.6 years, SD = 15.5.  
*Control group*: 31 participants (18 males and 13 females) Mean age = 41.3 years, SD = 14.5. | Intervention group: 20 (62.5%) participants with mild/moderate ID and 12 (37.5%) with severe/profound ID.  
Control group: 22 (71%) participants with mild/moderate ID and 9 (29%) with severe/profound ID. Diagnostic measures used were not identified. | Intervention group: Median ABC = 36  
Control group: Median ABC = 47  
Measured using the ABC | Behaviour therapy team (applied behaviour analysis and positive behavioural support) in addition to TAU | TAU (standard care) |
| **Ingersoll et al, 2016; USA, Pilot RCT** | Residential homes; 19 adolescents with ASD, severe ID and CB.  
*Intervention group*: 10 participants (9 males and 1 female) Mean age = 16.26 years, SD = 3.15.  
*Control group*: 9 participants (7 males 2 females) Mean age = 16.90 years, SD = 2.53. | ID level: severe;  
*Treatment group*: Mean IQ = 48.50 (SD = 13.42)  
*Control group*: Mean IQ = 47.43 (SD = 8.48)  
Measured using WAIS-IV. | Treatment group: Mean ABC = 43.70 (SD = 9.50)  
Control group: Mean ABC = 40.88 (SD = 6.92)  
Measured using the ABC-R. | Individual Reciprocal Imitation Training in addition to TAU (education, PBS plans, medication) | Individual sessions with teachers in addition to TAU |
| **Johnson et al, 2013; USA, Pilot RCT** | Inpatient; 33 children with ASD or PDD-NOS and sleep problems.  
*Intervention group*: 15 participants (11 boys and 4 girls) Mean age = 3.5 years, SD = 0.98.  
*Control group*: 18 participants (15 boys and 3 girls) Mean age = 3.6 years, SD = 1.12.  
Diagnostic measures used: ADI-R and ADOS. | Intervention group: 10 out of 15 children with lower cognitive scores (≤ 70 standard score) Mean standard cognitive score = 65.73, SD = 17.23.  
Control condition: 10 out of 18 children with lower cognitive scores (≤ 70 standard score) Mean standard cognitive score = 68.11, SD = 17.48.  
Measured using the MEIS scale or the SIIS scale. | All participants had sleep problems (e.g. bedtime resistance, night awakening) based on parents’ reports. | Behavioural parent training in addition to TAU (psychoeducational programme) | TAU (psychoeducational programme) |
<table>
<thead>
<tr>
<th>Study, location and design</th>
<th>Sample</th>
<th>ID level and its measure</th>
<th>CB and its measure</th>
<th>Intervention</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reitzel et al, 2013; Canada, Pilot RCT</strong></td>
<td>Autism Intervention Program; 15 children with ASD and early learning difficulties measured using the ELM.</td>
<td>Mean IQ = 37.1 (SD=12.4); mental age one of third below their chronological age. Scores were obtained using the Early Learning Measure.</td>
<td>Treatment group: Mean DBS score (CB measure) = 21.7 (SD=17.2). Control group: Mean DBS score = 21.4 (SD=24.7).</td>
<td>Functional Behaviour Skills Training in addition to TAU</td>
<td>Waitlist-control group</td>
</tr>
<tr>
<td><strong>Singh et al, 2016; USA, RCT</strong></td>
<td>Residential homes; 48 adults with severe or profound ID and CB.</td>
<td>Intervention group: 9 with severe ID and 25 with profound ID. Treatment group: 18 (75%) had behaviour plans for aggressive behaviours. Control group: 16 (67%) had behaviour plans for aggressive behaviours. Standardised measures to assess CB were not reported.</td>
<td>Mindfulness based PBS training (MBPBS)</td>
<td>PBS training</td>
<td></td>
</tr>
</tbody>
</table>

Note: RCT= Randomised Controlled Trial; ASD= Autistic Spectrum Disorder; ID= Intellectual Disabilities; CB= Challenging Behaviour; WAIS-IV= Wechsler Adult Intelligence Scale-Fourth Edition; ABC-R= Aberrant Behaviour Checklist-Residential; TAU= treatment as usual; PDD-NOS= Pervasive Developmental Disorder, not otherwise specified; MEIS= Multifactor Emotional Intelligence Scale; SIIS= Stanford–Binet Intelligence Scale; ADI-R= Autism Diagnostic Interview Revised Version; ADOS= Autism Diagnostic Observation Schedule; ELM= Early Learning Measure; DBS= Developmental Behavior Checklist – Parent/Carer Version.
Table 4: Elements of fidelity reported in the included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Adherence</th>
<th>Elements of Fidelity (Based on Dane and Schneider, 1998)</th>
<th>Responsiveness</th>
<th>Program differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hassiotis et al, 2009</td>
<td>Reported that fidelity was “high”. Methodological techniques used to measure fidelity not reported.</td>
<td>Frequency and duration: Mean contacts=9 (SD=7); Mean duration= 6 months.</td>
<td>Not reported.</td>
<td>Not reported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The intervention was provided by the behaviour therapy team.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of sessions completed: not reported.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantitative data: not reported.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingersoll et al, 2016</td>
<td>Six sessions randomly selected and assessed using fidelity checklists by the research team.</td>
<td>Frequency and duration: 10 min sessions 2 per day, 2–3 days per week.</td>
<td>Training (lasted 2 weeks) and supervision was provided for intervention providers by a supervisor who had previously established fidelity in RIT. RIT fidelity checklists were used to assess whether intervention providers learnt RIT and able to implement it.</td>
<td>Questionnaires administered to teachers to assess benefits and participants’ enjoyment of the intervention. One participant often wanted to finish sessions earlier which might indicate that he found the intervention unpleasant.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of sessions completed: not reported.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantitative data: not reported.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson et al, 2013</td>
<td>Used two fidelity measures: checklists and videotapes. Fidelity checklists were completed by therapists after each session. Therapists rated themselves on a scale 0 to 2 whether they achieved the goals of the intervention. The fidelity checklist also included assessment of parents’ adherence to the treatment (e.g. whether parents completed homework or implemented new strategies). All intervention sessions were videotaped and 10 % of them were randomly selected and assessed by independent observers.</td>
<td>Frequency and duration: five individual sessions (60-90 minutes).</td>
<td>Training was administered by two master-trained doctoral students or one senior doctoral-trained behaviour analyst. Training providers were regularly observed by an investigator and feedbacks were provided.</td>
<td>Satisfaction questionnaires were administered to parents to rate the usefulness of the teaching and specific elements of the intervention. Higher scores indicated greater satisfaction.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of sessions completed: self-reports indicated that the 15 families in the intervention group attended 73 of the 75 possible sessions (97.3%).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantitative data: the fidelity for the intervention=98% (range, 83–100) and parent adherence= 93% (range, 75–100).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantitative data: not reported.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Adherence</td>
<td>Dose</td>
<td>Quality of delivery</td>
<td>Responsiveness</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Reitzel et al, 2013</td>
<td>Staff completed fidelity checklists which were used to ensure protocols were followed and feedback was provided appropriately.</td>
<td>Frequency and duration: Two hour group sessions were delivered once per week for four months involving both children and their parents.</td>
<td>Training was provided by experienced therapists with post-secondary degrees or diplomas in ABA or IBI. Group sessions were staffed at a 1-to-1 child to staff ratio and were led by a teacher, a parent coach, and child prompters. Parents’ knowledge of ABA principles was assessed using a questionnaire developed for this study. Intervention group: Mean knowledge improvement after the training= 3 items (out of 15; SD = 2.5). Control group: Mean knowledge improvement after the training= 0.6 items (out of 15; SD = 2.5). Parents’ sense of competence was assessed using a questionnaire developed for this study. Intervention group: Mean= -11.7 points (SD = 15.8). Control group: Mean= -5.5 points (SD = 10.1).</td>
<td>Not reported.</td>
</tr>
<tr>
<td>Study</td>
<td>Adherence</td>
<td>Dose</td>
<td>Quality of delivery</td>
<td>Responsiveness</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Singh et al, 2016</td>
<td>Not reported.</td>
<td>Frequency and duration: On average, the caregivers meditated for 89% (range 0-96%) of the 40 days for between 25-40 minutes daily.</td>
<td>Training was presented in 3 parts over a 10 week period. A training provider was experienced behavioural analyst and mindfulness practitioner. Ten randomly selected training segments were videotaped and one observer (qualified in behaviour analysis and meditation) rated the fidelity of the training. <strong>Quantitative data:</strong> Mean fidelity rating for the training = 100%. Not reported.</td>
<td>Not reported.</td>
</tr>
</tbody>
</table>

Note: ICCs=intraclass correlation coefficients; RIT= Reciprocal Imitation Training; ABA=Applied Behaviour Analysis; IBI= Intensive Behaviour Intervention.
Supplementary material

Supplementary material 1: Search strategy

Ovid Medline, Embase and PsycINFO databases were searched together on 22 March 2017.

Web of Science and CINAHL Plus databases were searched separately on 15 June 2017 using the same search strategy and terms as with Medline, Embase and PsycINFO databases.

The following search terms were used:

1. “learning disab*”
2. “learning difficult*”
3. “learning disorder”
4. “learning impair*”
5. “intellectual* disab*”
6. “intellectual* impair*”
7. “intellectual dysfunction”
8. “development* disab*”
9. “development* disorder*”
10. “development* impair*”
11. “intellectual developmental disorder”
12. “mental* deficien*”
13. “mental* retard*”
14. “mental* handicap*”
15. “mental* disab*”
16. “mental insufficiency”
17. “mental* impair*”
18. “mental* challenged”
19. “intelligence”
20. IQ
21. “subaverage intelligence”
22. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22
23. “challenging behavio*”
24. “aberrant behavio*”
25. “maladaptive behavio*”
26. “problem behavio*”
27. “behavio* problems”
28. self-injury
29. self-harm
30. “self-injurious behavio*”
31. stereotypy
32. “stereotyped behavio*”
33. “repetitive behavio*”
34. aggress*
35. “destructive behavio*”
36. “property destruction”
37. “disruptive behavio*”
38. 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38
39. “positive behav* support”
40. PBS
41. “applied behav* analysis”
42. ABA
43. “functional analysis”
44. “functional assessment”
45. “behavio* intervention*”
46. “behavio* technique*”
47. “behavio* treatment*”
48. “behavio* support”
49. “behavio* modification”
50. 39 OR 40 OR 41 OR 42 OR 43 OR 44 OR 45 OR 46 OR 47 OR 48 OR 49
51. fidelity
52. treatment
53. intervention*
54. implement*
55. therap*
56. clinician
57. 52 OR 53 OR 54 OR 55 OR 56
58. 51 AND 57
59. Integrity
60. intervention*
61. treatment
62. program
63. 60 OR 61 OR 62
64. 59 AND 63
65. Adherence
66. differentiation
67. compliance
68. clinician
69. therap*
70. 65 OR 66 Or 67
71. 68 OR 69
72. 70 AND 71
73. 58 OR 64 OR 72
74. “randomised controlled trial”
75. “randomized controlled trial”
76. RCT
77. “random allocation”
78. randomisation
79. randomization
80. intervention
81. “clinical trial”
82. “control* trial*”
83. “randomly allocated”
84. “single blind*”
85. 74 OR 75 OR 76 OR 77 OR 78 OR 79 OR 80 OR 81 OR 82 OR 83 OR 84
86. 22 AND 38 AND 50 AND 73 AND 85
87. Limit 86 to English Language
88. Limit 87 to human
89. Limit 88 to year= “1990-Current”
90. Remove duplicates from 89
Supplementary material 2: Excluded studies

33 articles were excluded for the following reasons:

- Articles were not randomised control trials (N=8)
- Articles did not specify percentages of participants with intellectual disabilities in the sample (N=8)
- Interventions were based not on ABA or PBS principles (N=5)
- Interventions did not focus on reducing challenging behaviour (N=6)
- Results of fidelity measures have not been reported (N=1)
- More than 50% of sample had participants with IQ above 70 (N=3)

The following studies were excluded because they were not randomised control trials (N=8):


The following studies were excluded because percentages of participants with intellectual disabilities were not specified in the sample (N=8):


8. Thomeer, ML, Lopata, C, Volker, MA, Toomey, JA., Lee, GK. Audrey M. and Smith, RA. (2012). 'Randomized Clinical Trial Replication of a Psychosocial Treatment for

The following studies were excluded because interventions were not based on ABA or PBS principles (N=5):


The following studies were excluded because interventions did not aim to reduce challenging behaviour (N=6):


The following study was excluded because results of fidelity measures have not been reported (N=1):


The following studies were excluded because more than 50% of sample had IQ above 70 (N=3):


### Hassiotis et al, 2009

<table>
<thead>
<tr>
<th>Elements of risk of bias</th>
<th>Reviewers’ judgement</th>
<th>Description or Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence generation (allocation bias)</td>
<td>Low risk of bias</td>
<td>Participants were randomised to the groups using a computer-driven randomization list.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk of bias</td>
<td>Quote: “A set of sealed envelopes, each bearing only the name of the area and a number, were held by an independent administrator”.</td>
</tr>
<tr>
<td>Blinding of personnel and participants (performance bias)</td>
<td>Unclear risk of bias</td>
<td>No information provided.</td>
</tr>
<tr>
<td>Blinding of outcome assessors (detection bias) for all outcomes</td>
<td>Unclear risk of bias</td>
<td>No information provided.</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Low risk of bias</td>
<td>Quote: “One participant in each arm died during the trial, and one participant in the intervention arm declined to participate in follow-up assessments”.</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>High risk of bias</td>
<td>Stated that fidelity ratings were high but how it was measured and quantitative data were not provided.</td>
</tr>
<tr>
<td>Other potential threats to validity</td>
<td>High risk of bias</td>
<td>Moderate or larger treatment effects might be easier to achieve due to small sample.</td>
</tr>
</tbody>
</table>

### Ingersoll et al, 2016

<table>
<thead>
<tr>
<th>Elements of risk of bias</th>
<th>Reviewers’ judgement</th>
<th>Description or Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence generation (allocation bias)</td>
<td>Low risk of bias</td>
<td>Randomisation using a coin tossing procedure was reported in the paper.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>High risk of bias</td>
<td>Using the coin tossing method the study personnel might be able to predict to which group the next participant will be assigned.</td>
</tr>
<tr>
<td>Blinding of personnel and participants (performance bias)</td>
<td>High risk of bias</td>
<td>No blinding of participants and personnel.</td>
</tr>
</tbody>
</table>
| Blinding of outcome assessors (detection bias) for all outcomes | High risk of bias | Social skills and challenging behaviours were measured by a psychologist who was kept blind to group assignments. However, “the imitation measures for cohort 1 were administered by the teachers who worked with the adolescents and, for cohort 2, the third author who was a staff supervisor
at the centre and was not blind to group assignment”.

Incomplete outcome data (attrition bias) | Low risk of bias | Small number of drop outs was reported and its reasons were provided in the paper.

Selective reporting (reporting bias) | Low risk of bias | We believe that all outcomes were reported.

Other potential threats to validity | High risk of bias | Moderate or larger treatment effects might be easier to achieve due to small sample.

RIT sessions slightly longer than TAU sessions.

RIT sessions administered in addition to PBS which also targets communication as a part of a plan. Unclear if results are due to specific RIT techniques or differences within the quality of PBS plans.

Johnson et al, 2013

<table>
<thead>
<tr>
<th>Elements of risk of bias</th>
<th>Reviewers’ judgement</th>
<th>Description or Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence generation (allocation bias)</td>
<td>Unclear risk of bias</td>
<td>Participants were randomly assigned to the groups but specific methods employed to achieve randomisation were not specified.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk of bias</td>
<td>No information provided.</td>
</tr>
<tr>
<td>Blinding of personnel and participants (performance bias)</td>
<td>Unclear risk of bias</td>
<td>No information provided.</td>
</tr>
<tr>
<td>Blinding of outcome assessors (detection bias) for all outcomes</td>
<td>High risk of bias</td>
<td>Sleep problems were rated by parents who participated in the intervention group.</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>Low risk of bias</td>
<td>Small number of drop outs was reported and its reasons were provided in the paper.</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>Low risk of bias</td>
<td>All outcome data was reported in the paper.</td>
</tr>
<tr>
<td>Other potential threats to validity</td>
<td>High risk of bias</td>
<td>Moderate or larger treatment effects might be easier to achieve due to small sample.</td>
</tr>
</tbody>
</table>
### Reitzel et al., 2013

<table>
<thead>
<tr>
<th>Elements of risk of bias</th>
<th>Reviewers’ judgement</th>
<th>Description or Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence generation (allocation bias)</td>
<td>Low risk of bias</td>
<td>Participants were randomly assigned to treatment groups using a table of random numbers.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Low risk of bias</td>
<td>Quote: “Investigators and research staff were blind to participants’ group allocation”</td>
</tr>
<tr>
<td>Blinding of personnel and participants (performance bias)</td>
<td>Low risk of bias</td>
<td>Quote: “...therapists involved in administering the FBST were blind to the study’s objective and were not involved in assessing children’s pre- and post-treatment performance”</td>
</tr>
<tr>
<td>Blinding of outcome assessors (detection bias) for all outcomes</td>
<td>Low risk of bias</td>
<td>Research team was blind to participants’ group allocation.</td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td>High risk of bias</td>
<td>Numbers of drop-outs were quite large (5 participants in the control group and 6 participants in the treatment group).</td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td>High risk of bias</td>
<td>Reported that fidelity checklists were administered to the intervention providers but quantitative data were not provided.</td>
</tr>
<tr>
<td>Other potential threats to validity</td>
<td>High risk of bias</td>
<td>In the treatment group children had higher levels of ASD symptom severity, cognitive abilities and adaptive functioning than children in the control condition. In the control condition 3 children were also involved in Intensive Behavioural Intervention and this might biased the results.</td>
</tr>
</tbody>
</table>

### Singh et al., 2016

<table>
<thead>
<tr>
<th>Elements of risk of bias</th>
<th>Reviewers’ judgement</th>
<th>Description or Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence generation (allocation bias)</td>
<td>Low risk of bias</td>
<td>The authors used random number generation to assign participants to the groups.</td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td>Unclear risk of bias</td>
<td>No information provided.</td>
</tr>
<tr>
<td>Blinding of personnel and participants (performance bias)</td>
<td>Unclear risk of bias</td>
<td>No information provided.</td>
</tr>
<tr>
<td>Bias Type</td>
<td>Risk of Bias</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>----------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Blinding of outcome assessors</td>
<td>Unclear risk of bias</td>
<td>Reported that staff recorded outcome data but it is unclear if they were blinded to the treatment allocation.</td>
</tr>
<tr>
<td>Incomplete outcome data</td>
<td>Low risk of bias</td>
<td>Small number of drop outs was reported and its reasons were provided in the paper.</td>
</tr>
<tr>
<td>Selective reporting</td>
<td>Low risk of bias</td>
<td>We believe that all outcomes were reported.</td>
</tr>
<tr>
<td>Other potential threats to validity</td>
<td>High risk of bias</td>
<td>Several outcomes (aggression and physical restraints) were measured looking at records in the databases made by staff members. It is unclear whether staff members were blind to the treatment allocation and what steps were used to ensure that physical restraints were recorded accurately. Small sample size.</td>
</tr>
</tbody>
</table>