A Systematic Review of Animal-Assisted Therapy on Psychosocial Outcomes in People with Intellectual Disability

Sarah Maber-Aleksandrowicz a,*, Cerian Avent b,1, Angela Hassiotis c

a Kingston Community Team for People with Learning Disabilities, Hollyfield House, 22 Hollyfield Road, Surbiton KT5 9AL, UK
b Access Child and Youth Mental Health, Building 10, Springfield Hospital, 61 Glenburnie Road, London SW17 7DJ, UK
c Division of Psychiatry, Charles Bell House, 67-73 Riding House Street, London W1W 7EY, UK

A R T I C L E   I N F O

Article history:
Received 22 March 2015
Received in revised form 27 November 2015
Accepted 8 December 2015
Available online 7 January 2016

Keywords:
Animal-Assisted Therapy
Intellectual Disability
Psychosocial Outcomes

A B S T R A C T

The aim of the study was to review the literature on animal assisted therapy (AAT) in people with intellectual disabilities (ID) measuring psychosocial outcomes (behavioural, social, cognitive and emotional). Quantitative studies were found through a systematic search that identified studies using AAT in people with ID and measuring psychosocial outcomes (behavioural, cognitive, emotional and social). The quality of studies was assessed using a standardised tool and rated as strong, moderate or weak. Only published articles from peer-reviewed journals were included. No language or age restrictions were applied. Over half of the included studies were identified outside standard database searches (e.g. hand searching reference lists from included articles, references from AAT websites and using Google Scholar and a Grey Literature Database). Ten studies were included in the final review; two were rated as moderate quality and eight were rated as weak quality. Overall there was a positive improvement reported from studies for all psychosocial outcomes (with some cognitive, behavioural, social, emotional components reaching statistical significance p ≤ 0.01). Despite having no age restrictions, the included studies had participants that were mainly children and adolescents, in particular favouring male participants, which may limit generalisation. More rigorous methodology is required to improve the quality of future studies including in the main multicentre randomised designs and improved reporting according to CONSORT criteria. Further research should expand to include adults with ID and specific disorders such as challenging behaviour or mental illness.

© 2015 Elsevier Ltd. All rights reserved.

Contents

1. Method ................................................................. 324
  1.1. Protocol .......................................................... 324
  1.2. Eligibility Criteria ................................................. 324
  1.3. Information Sources ............................................. 325
  1.4. Search Strategy .................................................. 325

* Corresponding author. Tel.: +44 20 8949 9978.
E-mail addresses: sarahmaber@doctors.org.uk (S. Maber-Aleksandrowicz), cerian.avent@swlstg-tr.nhs.uk (C. Avent), a.hassiotis@ucl.ac.uk (A. Hassiotis).
1 Tel.: +44 20 3513 6609

http://dx.doi.org/10.1016/j.ridd.2015.12.005
0891-4222/© 2015 Elsevier Ltd. All rights reserved.
Animal-assisted therapy (AAT) is a form of therapy which aims to provide a therapeutic intervention for humans by involving animals in their treatment. Its focus is on augmenting behavioural, social, emotional, cognitive or physical functioning and is often a structured intervention with set goals and measured outcomes with precise definitions having recently been advanced by the International Association of Human-Animal Interaction Organizations (“IAHAIO White Paper,” 2014).

The aim of this review was to carry out a systematic review of AAT on psychosocial outcomes in people with Intellectual Disabilities (ID). To our knowledge there is no published systematic review yet regarding AAT specifically for people with intellectual disability (ID).

Animal-assisted therapy has been used throughout history to help people, with one of the first recorded uses of animals in a therapeutic purpose being in the York Retreat that opened in 1796 with a focus on rehabilitation for the mentally ill as opposed to the treatment in the asylums (Serpell, 2010). The use of AAT historically still relates to the rationale for AAT today that of the principle of rehabilitation and multidisciplinary approach to a person’s treatment. This is something which is especially pertinent when working with people with ID.

Studies have examined the impact of AAT and human-animal interaction on various aspects of physical health, mental health, in children with autism and in people with disabilities. One of the earliest meta-analysis investigating the effectiveness of AAT found effect sizes for changes in behavioural and medical problems to be in the moderate range, for well-being in the low to moderate range and for reducing autistic spectrum symptoms in the high range (Nimer and Lundahl, 2007). Whilst this meta-analysis attempted to group findings by age, presenting problem (medical, mental, behavioural) and participants’ functioning (normal or delayed), it did not identify participants with mental or physical disabilities. The authors noted whether participants had a “life-long disability” but this group included a mixed population of “autism, developmental delays, mental retardation or physical disabilities”. The authors considered 49 studies in their meta-analysis, but the disadvantage of such a large number is the loss of specificity of each study, particularly in being able to identity relevant effects in certain populations, such as those with intellectual disability. Furthermore, it was noted that there was considerable variance in the studies’ outcomes for individuals with disabilities compared to individuals with no disabilities; behavioural and well-being dependent variables included negative values in the confidence intervals for disabled individuals, whilst medical outcome dependent variables showed greater effect sizes for disabled individuals compared with individuals without disabilities.

A more recent meta-analysis (Virués-Ortega et al., 2012) of 21 studies examined the effects of AAT on psychological and functional status in populations with poor social functioning including elderly participants and those with depression and schizophrenia. Whilst it found moderate effects for AAT on depression, anxiety and behavioural disturbances it cautioned the interpretation of the findings due to inconsistent methodological characteristics of the included studies. The authors chose their selected populations as they represented “extensive at-risk groups for functional deterioration, low social support and social isolation” as there was evidence that individuals prone to these were likely to benefit from AAT (McNicholas and Collis, 2006). People with ID also fall into this category as they have limited social support structures (Lippold and Burns, 2009) and are at risk of social isolation (Jawaid et al., 2012). Therefore it follows that individuals with ID may also benefit from AAT interventions.

Another review examining AAT broadly across the literature was a systematic review (Matuszek, 2010) of animal-facilitated therapy in various populations and settings. This reported benefits for populations of hospitalised patients (in particular with heart failure, paediatric patients and those with pervasive developmental disorders), psychiatric patients, palliative care patients and also war veterans. There was also improvement seen when AAT was applied in various settings such as correctional facilities as well as in residential and nursing homes for the elderly. Whilst the positive effects of AAT were reported in all these populations and settings, the review suggested that further research was required specifically calling for larger sample sizes over longer periods of time with the addition of control groups and randomisation.
Two recent systematic reviews attempted to examine only randomised controlled trials (RCTs) on AAT in multiple populations (Kamioka et al., 2014) and (Maujean et al., 2015). The former review examined the effectiveness of AAT in 11 studies noting that it “may be an effective treatment for mental and behavioural disorders such as depression, schizophrenia and alcohol/drug additions”. However, due to poor methodological and reporting quality and heterogeneity there was insufficient evidence in the studies to offer clearly any conclusions about the effects of AAT based on RCTs. The later review examined psychosocial outcomes of AAT in 7 studies noting that all but one study reported at least some beneficial psychological effects including: improved self-motivation, self-efficacy, self-esteem and decreased depression, anxiety, positive and negative symptoms of schizophrenia compared to control groups. However, again there were many concerns raised regarding the methodology of the included studies, specifically that there was substantial variability in regard to quality and in the measures used to assess psychological outcomes. The authors concluded that whilst there was evidence that AAT may be beneficial in a wide range of individuals, different effects may be found depending on the population investigated and emphasised the need to assess the efficacy of AAT interventions in a wide range of populations. They called for further research to understand what specific types of AAT would be beneficial in specific populations. This is also echoed by another systematic review (Bernabei et al., 2013) examining the effects of AAT on elderly patients with dementia or various psychiatric disorders which found positive influences on patients with dementia by reducing agitation, improving social interaction as well as communication and coping ability whilst the few studies on elderly patients with psychiatric diagnoses produced more mixed findings concluding more research was required in the duration and frequency of sessions as well as a suitable target group. Hence the purpose of this review therefore was to target a specific population, i.e. ID, in examining AAT on psychosocial outcomes.

Previous reviews of AAT studies that have focused exclusively on a sole condition or population include dementia (Filan and Llewellyn-Jones, 2006), depression (Souter and Miller, 2007) and autistic spectrum disorder (O’Haire, 2013). The systematic review of AAT in people with dementia examined 11 studies and concluded that AAT improved behavioural and psychological symptoms in this group but noted that the duration of the effect was not explored as well as noting the positive effect on staff could be a confounder. The meta-analysis of five studies examining AAT’s effect on depression supported a therapeutic benefit of using dogs for people with depression with effects that were “both statistically significant and large enough to be of practical significance”. The systematic review evaluating animal-assisted intervention for autism spectrum disorder (ASD) (O’Haire, 2013) comprised 14 studies investigating several outcomes such as social interaction, language and communication, ASD severity, problem behaviours and stress and well-being. The studies reported positive outcomes but had significant bias including small sample sizes, lack of control groups and poor ascertainment process which limit generalisation. The latter is particularly important to this review, as there is evidence to suggest higher initial cognitive ability may be correlated to better progress in therapy outcome in children with ASD (Ben-Itzchak and Zachor, 2007).

Scrutiny of methodology is important for all reviews of intervention and outcomes. In particular, the rigor and quality of AAT studies has sometimes been heavily criticized for being subject to bias such as non-specific treatment effects, construct confounding (Marino and Lilienfeld, 2007) as well as novelty effects (Marino, 2012). A common theme concluding from the meta-analyses and systematic reviews mentioned above is that research in the field of AAT continues to have various and significant methodological weaknesses and minimal standards of research design. Many of the meta-analyses and systematic reviews of AAT have also excluded non-English published papers which may have potentially narrowed the field further and created bias.

1. Method

1.1. Protocol

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist (2009) was used for this systematic review based on the PRISMA statement guidelines (Moher et al., 2009). The PRISMA statement aims to improve reporting in systematic reviews and meta-analyses by setting out a minimum standard of evidence-based reporting and consists of a four-phase flow diagram and 27-item checklist.

1.2. Eligibility Criteria

The following inclusion criteria were applied to the review:

**Participants.** People with ID of any age. This was defined by studies naming participants as having intellectual disability or mental retardation (the latter reflecting the older medical definition of ID and World Health Organisation Definition in the ICD-10). For other terms such as Learning Disability, Mental Handicap, Mental Disability if there was an accompanying reported IQ < 70 for participants (the ICD-10 cut off for people with ID), the selected study was eligible. These latter terms may reflect language of the time or country of origin. Terms such as Learning Disabilities have been confused with Learning Difficulties such as specific reading disorders, hence why the addition of IQ measurement was required for such studies to be included. Participants with genetic conditions which always result in ID as part of the phenotype were eligible e.g. Down’s syndrome, Angelman Syndrome. Participants with ASD were eligible only if they also had ID. Studies with groups of mixed diagnosis participants (i.e. ID, dementia, cerebral palsy) were eligible if > 85% had ID identified (by number/percentage) or if the authors provided subgroup analysis for the ID group.
Intervention. Studies which incorporated live animals as part of therapy, as defined by the European Society for Animal-Assisted Therapy (ESAAT) website:

“Animal-assisted therapy includes deliberately planned pedagogic, psychological and socially integrative interventions with animals for children, youths, adults and senior citizens with cognitive, social-emotional and motoric disabilities, and behavioural problems, and for focused support. It also includes health-promoting, preventive and rehabilitative measures. Animal-assisted therapy takes place individually and within a group setting. Animal-assisted therapy is based on the relationship and process structure within a triangular relationship between the client, animal and therapist. Animal-assisted therapy involves methods by which clients interact with animals, communicate via animals or are active for animals. Implementation is goal-oriented and based on a clear process and topic orientation taking into account animal-ethical principles with subsequent documentation and professional well-founded considerations.” (ESAAT, Definition Animal Assisted Therapy, (n.d.)).

Outcomes. In this study, psychosocial outcomes corresponded to reports of cognitive, social, emotional or behavioural functioning in the literature and measured by observer rated or self-reported scales. For the purpose of this review, cognitive outcomes corresponded to functions relating to memory, concentration, focus/attention or knowledge. Social outcomes corresponded to interpersonal relationships such as between the subjects and their peers, or therapists, or facilitators, or observers, or therapeutic animals, as well as the subjects' communication with others (verbal or non-verbal). Emotional outcomes corresponded to the subject's general mood or feelings such as happiness, sadness or anxiousness. Behavioural outcomes corresponded to the general conduct of the subject such as aggression, hyperactivity, inappropriate words/acts as well as levels of independent/autonomous behaviour.

Study Type. Quantitative studies (i.e. those measuring outcomes using numerical scales) collecting primary data. These included randomised controlled trials, quasi-randomised controlled trials, controlled and uncontrolled before and after studies and quantitative case studies including single-subject research design. The studies were primary research published in peer reviewed journals without date or language restrictions.

The inclusion criteria were based upon the Cochrane Participants, Interventions, Comparisons, Outcomes and Study Type (PICO(S) acronym) from the online Cochrane Handbook for Systematic Review Interventions (Higgins and Green, 2011). Whilst the inclusion criteria may have appeared strict on our intended population with regards to diagnosis (ID), to balance this stringency, wider criteria were applied to other areas namely no age or participant number restrictions, no exclusions on type of quantitative study design, nor on type of measurements for psychosocial outcomes. This was to obtain collectively as many studies as possible examining AAT on psychosocial outcomes specifically in the ID population. Moreover, studies which came close to meeting the inclusion criteria but failed due to mixed patient populations (including ID participants) were collected and have been summarised in the Appendix Table A1.

1.3. Information Sources

A total of six databases were searched from their start date until December 2012: Web of Knowledge (including Web of Science 1900-Present and Medline 1950-Present), PsycINFO (1806-Present), PubMed, CINAHL Plus (1991-Present), EMBASE (1980 – Present) and the Cochrane Library (Database of Abstracts of Reviews of Effects and Central Controlled Trials Register). The Grey Literature was searched using Google Scholar and System for Information on Grey Literature in Europe (SIGLE) website (http://www.opengrey.eu/). Also the reference lists from included articles were hand searched. Furthermore the website of a prominent AAT society, Pet Partners (http://www.petpartners.org), previously known as the Delta Society, was searched for extra references. These searches were then re-run in December 2014 but no new studies were included. One study author (Klimberg, 2002) was contacted to clarify the measurement tools used in her study. It was subsequently not included as the outcomes were not recorded using a quantitative method.

1.4. Search Strategy

Search terms for ID were combined with search terms for AAT. The terms were searched in the title without date or language restriction. The Web of Knowledge search was conducted as follows:

learning disabilit* OR learning difficul* OR intellectual disabilit* OR mental retardation OR mental handicap* OR IQ OR Down* Syndrome OR autis* OR intellectual impairment OR challenging behavio* OR behavio*
AND
animal-assisted therap* OR animal assisted therap* OR pet-facilitated therap* OR pet facilitated therap* OR pet-assisted therap* OR pet assisted therap* OR therap* OR kynotherapy OR dolphin* therap* OR horse* therap* OR hippotherapy OR animal-assisted activit* OR animal assisted activit*

1.5. Study Selection

The first author (SMA) conducted the initial search in December 2012 as well as the second search in December 2014. She scanned all the titles from the identified studies, then read the relevant abstracts, retrieved studies and applied the eligibility criteria to each retrieved paper which led to final papers for inclusion in the systematic review. Using a standard form, details of the study including its participant characteristics, intervention description – type and duration, as well as study outcomes
were also recorded. Additional information such as year of publication, name of journal, country study was conducted in, country of author and language of publication were also noted as data which could be examined.

1.6. Data Collection Process and Reporting of Study Results

Data were extracted from the included studies using a standardised data abstraction form, from the Effective Public Health Practice Project’s (EPHPP) website (EPHPP, Quality Assessment Tool for Quantitative Studies, 2009) which examines both internal and external validity. Using the tool and accompanying dictionary it allowed studies to be rated as strong, moderate or weak in the following criteria: selection bias, allocation bias (study design), confounders, blinding, data collection methods and withdrawals and drop-outs. It also further has comments on intervention integrity and appropriate statistical analysis. A final global rating is then allocated for the individual paper depending on the number of weak ratings in the examined six components. A global rating of strong is given if there are no weak component ratings, a global rating of moderate is given if there is only one weak component rating, and a global rating of weak is given if there are two or more weak component ratings.

The first (SMA) and second (CA) authors independently read the ten included papers scoring them using the Quality Assessment Tool for Quantitative Studies. Both reviewers then discussed the ratings and where there were discrepancies between the two reviewers with respect to the six component ratings a joint review and decision was taken and a final global rating allocated. Although not a part of the standardised tool, the reviewers additionally allocated numbers for each of the six rated components; weak ratings allocated one, moderate ratings allocated two and strong ratings allocated three. This provided a cumulative score, allowing ranking of the papers according to methodological rigor in a tabular form, alongside their standardised global rating. The principal summary measures gave a narrative indication of the most frequently reported improvement or deterioration in psychosocial outcomes for each study. These are recorded in Table 1.

2. Results

2.1. Study Selection

The search resulted in 2750 (after duplicates removal) studies whose titles and abstracts were screened. Fig. 1 shows the study selection process by a flow diagram for the search. A significant proportion of the excluded papers (over half of all studies identified,) at the abstract screening stage, were neuro-imaging studies using PET (Positron Emission Tomography). Other excluded papers at this stage included genetic/animal model studies, veterinarian studies and zoonotic diseases as well as papers not including any animals or having any participants with ID. This left 47 articles which were assessed eligible based on their abstracts and were reviewed in full. Of the retrieved full text articles 36 were then excluded, many due to having mixed patient populations without subgroup analysis for the ID group, not a clear number or percentage of people with ID, or participants with ID < 85% of population. These excluded studies are reported in more detail in Appendix Table A1 which found similar results to the systematic review looking at animal-assisted intervention in ASD populations (O’Haire, 2013) containing mostly studies related to dolphin-therapy. One further study (Klimberg, 2002) was excluded after contacting the study’s author to clarify the outcome measures used in her study. These were confirmed to be qualitative methods (parents’ opinions) and hence fell outside of the inclusion criteria of this systematic review. Ten studies were included in the final review.

2.2. Study Characteristics

The ten studies included in the review are presented in Table 1. These show an overview of the study characteristics, information on individual study participants, individual study interventions, main outcomes and assessment of the quality rating of the papers.

2.3. Included Studies

Dates of Publication and Country of Study. The 10 studies were published in the years 1989 – 2012, with over half from the last decade. Four studies came from the USA and the rest from Bosnia & Herzegovina, Czech Republic, Italy, Poland, Slovakia, and UK. Of the full articles for retrieval in the December 2012 search, a total of eight (8 of 47, 17%) were not published in English (4 French, 2 German, 1 Polish, 1 Czech). One of the final ten studies included in the review was not an English publication (this was a Czech paper that the first author, SMA, subsequently had translated).

Data Sources of Study and Journal Publication. Of the included studies: five had been found through databases, three (Kršková et al., 2010; Memishević and Hodžičk, 2010; Cawley et al., 1994) were identified through hand searching references, two (Walters Esteves and Stokes, 2008; Borioni et al., 2012) from a website about AAT called Pet Partners (http://www.petpartners.org) and one (Heimlich, 2001) was found through the grey literature using Google Scholar and Wikipedia. The topic of journals in which they were published included: rehabilitation (3), psychology or neuropsychology (2), human-animal interaction (4) and autism (1).
2.4. Participants

**Age and Gender.** Children and adolescents were the participants in the majority of studies comprising 9 of the 10 studies, with only one study having an adult only population (Borioni et al., 2012). Also, most of the gender distributions favoured male participants with 7 of the 10 studies having > 50% male participants.
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Journal</th>
<th>Therapy terminology (Animal intervention)</th>
<th>Diagnosis description</th>
<th>Participants: Age (years) Gender</th>
<th>Intervention: Location, administrator &amp; format (group or individual)</th>
<th>Study design &amp; duration</th>
<th>Main outcomes &amp; measurement tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borioni et al., 2012</td>
<td>Disability &amp; Rehabilitation</td>
<td>Equestrian Rehabilitation &amp; Onotherapy (Horse &amp; Donkey)</td>
<td>Intellectual Disability</td>
<td>Horse Group: 23 (100%), Horse Group: 42.9 (mean), Horse Group: SD 1.5, Horse Group: M 7 (88%), Horse Group: F 1 (12%)</td>
<td>Donkey Group: 15 (100%), Donkey Group: 38.6 (mean), Donkey Group: SD 8.6, Donkey Group: M 13 (88%), Donkey Group: F 2 (12%)</td>
<td>Design: BA/ITS, Duration: Horse Group: 18 months, Donkey Group: 12 months, Session N &amp; length not specified</td>
<td>Location: Rehabilitation centre specialising in equine therapy, Administrator: Psychologist &amp; Equine Instructors, Format: Not specified/clear</td>
</tr>
<tr>
<td>Kříková et al., 2010</td>
<td>Society and Animals</td>
<td>Therapeutic Animal &amp; Animal-assisted therapy (Guinea Pig)</td>
<td>Intellectual Disability</td>
<td>Concurrent Diagnoses: 3 DS, 3 CP, 1A.</td>
<td>Location: Classroom, Primary School</td>
<td>Design: BA, Duration: 10 weeks, Once a week, 4 hours, 10 sessions</td>
<td>Measurement Tool: Frequency of social behaviours recorded by observer</td>
</tr>
<tr>
<td>Memishević and Hodžić, 2010</td>
<td>Journal of Special Education and Rehabilitation Bosnian &amp; Herzegovina</td>
<td>Equine-assisted therapy (Horse)</td>
<td>Intellectual Disability</td>
<td>Concurrent Diagnosis: A</td>
<td>Location: Riding centre</td>
<td>Design: BA, Duration: 10 weeks, Once a week, 30 minutes, 10 sessions</td>
<td>Measurement Tool: Autism Treatment Evaluation Checklists (ATEC)</td>
</tr>
</tbody>
</table>

Outcome: Improvement compared to baseline: in autonomy, affective-relational areas, mental cognitive areas & neuro-psychological areas**, Donkey Group Only: Improvement compared to baseline in communication*. **Therapy terminology (Animal intervention): Equestrian Rehabilitation & Onotherapy (Horse & Donkey), Kynotherapy (Dog), Therapeutic Animal & Animal-assisted therapy (Guinea Pig), Equine-assisted therapy (Horse).
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Species</th>
<th>Diagnosis</th>
<th>Age</th>
<th>Gender</th>
<th>Location</th>
<th>Administrator</th>
<th>Format</th>
<th>Design</th>
<th>Duration</th>
<th>Measurement Tool</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walters Esteves and Stokes, 2008 Anthrozooês USA (English)</td>
<td>Pet Therapy (Dog)</td>
<td>Mental Retardation</td>
<td>Children 5-9</td>
<td>Location: School classroom</td>
<td>Format: Individual</td>
<td>Design: BA/ITS</td>
<td>Duration: 5 sessions a week, 8 minutes</td>
<td>Measurement Tool: Social behaviours observed and categorised as positive/negative, verbal/non-verbal and initiations/responses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heimlich, 2001 Journal of Rehabilitation USA (English)</td>
<td>Animal-Assisted Therapy (Dog)</td>
<td>Mental Retardation</td>
<td>Children, Adolescents &amp; Young Adults 7-19</td>
<td>Location: Residential facility for children with multiple disabilities</td>
<td>Format: Individual</td>
<td>Design: BA</td>
<td>Duration: 8 weeks. 2 sessions a week, 30 minutes</td>
<td>Measurement Tools: Measurement of Pet Intervention (MOP)®, Direct Observation Form (DOF), Teacher's Report Form (TRF) of the Child's Behavior Checklist (CBC) &amp; Behavior Dimensions Rating Scale (BDRS)®.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limond et al., 2007 Anthrozooês UK (English)</td>
<td>Therapy Dog (Dog)</td>
<td>Learning Disability and Down’s Syndrome</td>
<td>Children 6-12</td>
<td>Location: School for children with severe learning difficulties</td>
<td>Format: Individual</td>
<td>Design: AT</td>
<td>Duration: 6 weeks. Once weekly for 6 sessions, 7 mins with artificial dog, 7 mins with real dog in the same session.</td>
<td>Measurement Tool: Duration &amp; frequency of social behaviours recorded.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chamrádová, 1995 Ceskoslovenska Psychologie Czech Republic (Czech)</td>
<td>Animal-assisted therapy (Dog)</td>
<td>Severe Mental Retardation</td>
<td>Children Age not specified</td>
<td>Location: Clinic</td>
<td>Format: Not specified/clear</td>
<td>Design: CCT Experiment group (exposed to real dog &amp; artificial dog). Control group (only exposed to toy dogs).</td>
<td>Measurement Tool: Frequency of 51 variables of behaviour (which author defined but did not describe in study) and duration of focused attention.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Notes:**
- **Incidence:** 3 (100%) Concurrent Diagnoses: 2 DS, 1 HI
- **Incidence:** 14 (100%): 1 (7%) MOD, 1 (7%) MOD - SEV, 8 (57%) SEV, 4 (29%) SEV - PR
- **Incidence:** 8 (100%) Concurrent Diagnoses: 3 CP, 3 SeD, 3 NV, 2 SpD, 2 Ast, 2 A, 1 DS
- **Incidence:** Learning Disability and Down’s Syndrome 8 (100%) Concurrent Diagnoses: 8 (25%) F 6 (75%)
- **Incidence:** Children Age not specified Gender recorded by author but not specified in paper
- **Incidence:** Severe Mental Retardation 16 (100%) Concurrent Diagnoses: Gender recorded by author but not specified in paper
- **Incidence:** Mental Retardation 3 (100%) Concurrent Diagnoses: 2 DS, 1 HI
- **Incidence:** Mental Retardation 14 (100%): 1 (7%) MOD, 1 (7%) MOD - SEV, 8 (57%) SEV, 4 (29%) SEV - PR
- **Incidence:** Children, Adolescents & Young Adults 7-19
- **Incidence:** Learning Disability and Down’s Syndrome 8 (100%)
- **Incidence:** Mental Retardation 3 (100%)
- **Incidence:** Mental Retardation 14 (100%): 1 (7%) MOD, 1 (7%) MOD - SEV, 8 (57%) SEV, 4 (29%) SEV - PR
- **Incidence:** Children, Adolescents & Young Adults 7-19
- **Incidence:** Learning Disability and Down’s Syndrome 8 (100%) Concurrent Diagnoses: 8 (25%) F 6 (75%)
- **Incidence:** Children Age not specified Gender recorded by author but not specified in paper
- **Incidence:** Severe Mental Retardation 16 (100%) Concurrent Diagnoses: Gender recorded by author but not specified in paper
- **Incidence:** Mental Retardation 3 (100%)
- **Incidence:** Mental Retardation 14 (100%): 1 (7%) MOD, 1 (7%) MOD - SEV, 8 (57%) SEV, 4 (29%) SEV - PR
- **Incidence:** Children, Adolescents & Young Adults 7-19
<table>
<thead>
<tr>
<th>Study reference</th>
<th>Journal</th>
<th>Country of study</th>
<th>(Publication language)</th>
<th>Therapy terminology</th>
<th>Diagnosis description</th>
<th>Participants:</th>
<th>Intervention:</th>
<th>Study design &amp; duration</th>
<th>Main outcomes &amp; measurement tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cawley et al., 1994</td>
<td>Anthrozoo¨s</td>
<td>USA</td>
<td>(English)</td>
<td>Therapeutic Riding (Horse)</td>
<td>Mental Retardation</td>
<td>4 of 23 (18%) Rest of Group: LD, SEH M 13 (57%) F 10 (43%)</td>
<td>Location: Riding centre Administrator: Volunteers Format: Group</td>
<td>Design: BA Duration: 8 weeks, 8 sessions, 2 hours</td>
<td>Measurement Tool: Piers Harris Children’s Self-Concept Scale&lt;sup&gt;g&lt;/sup&gt; Outcome: Whole group: Increase in self-concept compared to baseline. Subgroup analysis for MR group: Improved most on behaviour &amp; anxiety scores compared to baseline. However, both total &amp; sub-group scores did not reach statistical significance (no p value reported).</td>
</tr>
<tr>
<td>Redefer and Goodman, 1989</td>
<td>Journal of Autism and Developmental Disorders</td>
<td>USA</td>
<td>(English)</td>
<td>Pet-Facilitated Therapy (Dog)</td>
<td>Mental Retardation</td>
<td>12 (100%) Concurrent Diagnosis: AF</td>
<td>Location: Not reported/clear Administrator: Therapist Format: Individual</td>
<td>Design: BA Duration: 18 sessions, 20 minutes. Duration of intervention not reported</td>
<td>Measurement Tool: Social behavioural categories observed &amp; coded as isolation or social interaction&lt;sup&gt;h&lt;/sup&gt; Outcome: Improvement compared to baseline: decreased social isolation &amp; increased social interaction sustained over course of therapy**. Improvement compared to baseline at post-treatment*** &amp; by one-month follow up&lt;sup&gt;u&lt;/sup&gt; (both with no dog present) but eroding improvement from treatment to follow-up.</td>
</tr>
</tbody>
</table>

Note. A - Autism, AF - Autistic Features, AS - Asperger's Syndrome, Ast - Asthma, AT - Alternating treatment, BA - Before and after, CCT - Controlled Clinical Trial, CP - Cerebral Palsy, DS - Down's Syndrome, F - Female, HI - Hearing Impairment, ID - Intellectual disability, ITS - Interrupted Time Series, LD - Learning Disabled, M - Male, MOD - Moderate, MR - Mental retardation, NV - Non-verbal, PR - Profound, SD - Standard Deviation, SeD - Seizure Disorder, SEH - Severe Emotional Handicap, SEV - Severe, SpD - Spastic Diplegia. <sup>a</sup>Scale of Observation of the Behavior of Children and Parents - consists of 4 sub-scales assessing the development of: cognitive, emotional, social and motor functions. <sup>b</sup>Social behaviour represented by the sum of tactile, verbal and eye contacts (tactile contact = frequency of touch contacts towards other people or the animal, verbal contact = frequency of sounds and words addressed to people or the animal, eye contact = frequency of all variants of eye contacts aimed towards other people or the animal). <sup>c</sup>Positive verbal defined as utterances indicating pleasure or interest in the situation. Negative verbal defined as utterances of displeasure or disinterest in the situation. Positive non-verbal behaviours defined as behaviours indicating pleasure or interest in the situation. Negative non-verbal behaviours defined as behaviours indicating displeasure or disinterest in the situation. Interactions recorded as child-initiated toward the teacher or dog or teacher-prompted interactions toward the teacher or dog. <sup>d</sup>MOPI – consists of 4 items: attention span, physical movement, communication and compliance on a 7-point Likert scale. <sup>e</sup>BDRS - this has 4 subscales: aggressive/acting out, irresponsible/inattentive, socially withdrawn and fearful/anxious. Social behaviours were measured against the object at which they were directed and distinguished into 3 categories: looking behaviours, responding to the adult and initiating behaviours. Responding to the adult and initiating behaviours could be verbal or non-verbal, looking and initiating behaviours were directed at the test dog, handler, toys or coded as ‘other’ – including room fittings or undirected. Responding to the adult behaviours were directed at the handler. <sup>f</sup>Piers Harris Children's Self-Concept Scale consists of 6 subcategories: Behaviour, Intellectual and school status, Physical appearance and attributes, Anxiety, Popularity and Happiness and satisfaction. Social behavioural categories observed and coded as isolation (play or self-stimulatory activities directed to self) or social interaction (any instance of verbal or nonverbal behaviour directed to therapist or dog). *p ≤ .05. ** p ≤ .01. *** p ≤ .001.
Number of Participants per Study. The number of participants in the individual studies was overall low; with three studies having less than five participants. Three studies each had between 5-10, and 10-20 participants respectively, with only one study having above 20 participants (n = 23) (Borioni et al., 2012). There were a total number of 100 participants with ID across all studies.

Diagnosis of ID. In eight out of the ten studies all participants had ID, with one study having a participant population composed of 8% of people with ID (Krišková et al., 2010), and another one study with an 18% ID proportion of the sample population (Cawley et al., 1994). The later was included as data from this group was examined within a subgroup analysis. Four studies either described or tabulated the severity of ID of the sample as mild, moderate, severe or profound (Borioni et al., 2012; Memisevikj and Hodzhikj, 2010; Heimlich, 2001; Chamrádová, 1995). Eight studies reported participants having a concurrent diagnosis of autism, autistic features or Asperger Syndrome (5 studies) or Down Syndrome (3 studies). The terminology used to describe the participants ID diagnosis varied across time and countries. Four of the ten studies used the term ID in papers from 2010-2012, all from European studies. The term mental retardation was still the most prevalent, used in 5 of the 10 studies, and was used in American papers from 1989 – 2008 and in one European study in 1995. One paper from the UK described the population as having Learning Disability and Down’s Syndrome (Limond et al., 2007).

2.5. Intervention

Type of Therapy Animals. Out of the therapy animals used in the included studies, the dog was the most frequently used in six studies, then a horse (three studies), the only other two animals being a donkey and a guinea-pig in one study each respectively. Other animals used in the literature are shown in Appendix Table A1.

Therapy Terminology. The term used to define the therapy also varied widely with almost all studies using different terminology. AAT was the only term mentioned in more than one study (three studies in total), with the rest being listed in Table 1.

Intervention Location and Administrator of Therapy. The location of the intervention and the person giving the therapy continued to show diversity. School classrooms were the main location for the intervention (four studies) followed by riding centres (two studies). The person giving the therapy was most often a teacher or a therapist (type of therapist not defined) each in two included studies respectively.

Group or Individual Therapy. The intervention was given to individual participants in half of the studies (five studies). In two studies the intervention was given to groups and in the remaining three studies it was not specified.

Duration of Total Therapy Intervention. The majority of the studies (6 studies) had a duration of intervention lasting between 6-10 weeks, with two studies having a duration of 12 months or over. The remaining two papers did not report duration.

Duration of an Individual Therapy Session. The most common duration for an individual session was between 20-45 minutes (4 studies). This was followed by 10 minutes or less, 2-4 hours, or length of session not specified in two papers each respectively.

Design of Study Intervention. The designs adopted were either pre and post intervention studies (also known as cohort/one group pre and post) or before and after studies with multiple testing (eight studies). One study used an alternating treatment (artificial dog and real dog in the same session) (Limond et al., 2007) and the other one study was a controlled clinical trial but without randomisation (Chamrádová, 1995).

2.6. Outcomes

The individual study outcomes are described in Table 1. The majority of studies looked at multiple components of psychosocial outcomes, but for summarisation purposes, this will be categorised into social outcomes, cognitive outcomes, emotional outcomes and behavioural outcomes.

Social Outcomes. Nine of the ten studies examining social outcomes found that participants that had the animal interventions had increased social interaction/socialisation and an improvement in communication (both verbal and non-verbal). Three studies specifically reported an outcome of increased social interaction with p value < 0.01 (Krišková et al., 2010; Limond et al., 2007; Redefer and Goodman, 1989).

Cognitive Outcomes. Seven of the ten studies examining cognitive outcomes found that participants that had the animal intervention showed improvement in cognitive functions, concentration or attention. Three studies reported accompanying p values of < 0.01 for mental cognitive improvement (Borioni et al., 2012), improved concentration and in learning new tasks (Pawlik-Popielarska, 2010) and improved focus/attention (Limond et al., 2007).

Emotional Outcomes. Two of the ten studies found participants that had the animal intervention had a positive impact on their emotional well-being. One study reported significant improvements compared to baseline in mood and ability to relax with a p value ≤ 0.01 (Pawlik-Popielarska, 2010) whilst another reported improvements in anxiety scores compared to baseline but that it did not reach statistical significance (Cawley et al., 1994).

Behaviour Outcomes. Four of the ten studies found participants that had the animal intervention showed improvement in behaviour. The only study to report results with associated significant p values in behavioural outcomes was Borioni et al., 2012, which showed participants improved in autonomous behaviour (p < 0.01). Another study showed mixed results; improvement in behaviour (up to 11%) for female participants with mild ID, yet lower improvement (up to 4%) and in some
<table>
<thead>
<tr>
<th>Study Reference</th>
<th>First Author (SMA) Global Rating</th>
<th>Global Rating Consensus Both Reviewers</th>
<th>Second Author (CA) Global Rating</th>
<th>Global &amp; Component Rating Consensus Both Reviewers</th>
<th>Main Study Weakness</th>
<th>Reason for Discrepancy</th>
<th>Strength of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borioni et al. (2012)</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 8</td>
<td>Weakness: 1, 3, 5 &amp; 6</td>
<td>Agreed that all participants completed the pre &amp; post intervention scores so this was an acceptable measure of drop outs.</td>
<td>Largest sample size (23 participants) of included studies. Only study investigating adult ID population. Reported severity of ID.</td>
</tr>
<tr>
<td>Pawlik-Popielarska (2010)</td>
<td>Global: Weak Component: 9</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Weakness: 1, 3 &amp; 5</td>
<td>Agreed that all participants completed the pre &amp; post intervention scores so this was an acceptable measure of drop outs.</td>
<td>Tabulation of the animal intervention program giving a clear view of the treatment approach Detailed account of participants’ diagnoses Use of p values</td>
</tr>
<tr>
<td>Kršková et al. (2010)</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Weakness: 1, 3 &amp; 5</td>
<td>Agreed that all participants completed the pre &amp; post intervention scores so this was an acceptable measure of drop outs.</td>
<td>Detailed account of participants’ diagnoses Use of p values Data presented individually as well as for the group. Use of standardised measuring tools Reporting of adverse outcome &amp; participant withdrawal/drop-out. Valid and reliable data collection tool</td>
</tr>
<tr>
<td>Memishievikj and Hodzhikj (2010)</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Global: Weak Component: 10</td>
<td>Weakness: 1, 3 &amp; 4</td>
<td>Agreed that all participants completed the pre &amp; post intervention scores so this was an acceptable measure of drop outs.</td>
<td>Detailed account of participants. Use of some standardised measuring tools Reporting of adverse outcome &amp; participant withdrawal/drop-out. Valid and reliable data collection tool</td>
</tr>
<tr>
<td>Heimlich (2001)</td>
<td>Global: Moderate Component: 12</td>
<td>Global: Weak Component: 9</td>
<td>Global: Moderate Component: 12</td>
<td>Global: Moderate Component: 12</td>
<td>Weakness: 3</td>
<td>Agreed that all participants completed the pre &amp; post intervention scores so this was an acceptable measure of drop outs.</td>
<td>Valid and reliable data collection tool</td>
</tr>
<tr>
<td>Limond et al. (2007)</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 8</td>
<td>Global: Weak Component: 8</td>
<td>Weakness: 1, 3, 5 &amp; 6</td>
<td>Agreed the method for selection of participants was not clear.</td>
<td>Reported pet ownership for participants (confounder)</td>
</tr>
<tr>
<td>Cawley et al. (1994)</td>
<td>Global: Weak Component: 11</td>
<td>Global: Weak Component: 12</td>
<td>Global: Moderate Component: 13</td>
<td>Global: Moderate Component: 13</td>
<td>Weakness: 3</td>
<td>Agreed that all participants completed the pre &amp; post intervention scores so this was an acceptable measure of drop outs.</td>
<td>Examined the relationship between pet ownership &amp; the participant’s response to the animal in therapy (confounder). Reported adverse events</td>
</tr>
</tbody>
</table>

Note. Key to Quality Assessment: *Global Rating: Weak = two or more Weak component ratings, Moderate = one Weak component rating, Strong = no Weak component ratings. Component Ratings: Weak = 1, Moderate = 2, Strong = 3 giving total cumulative score of 18 for the 6 individual components. Individual Components: 1 = Selection Bias, 2 = Allocation Bias/Study Design, 3 = Confounders, 4 = Blinding, 5 = Data Collection Methods, 6 = Withdrawals and Drop-Outs.
cases deterioration in behaviour for males with severe ID (Memishevskij and Hodzhikj, 2010). In Heimlich, 2001, the study was terminated early, so analysis could not be completed, but the author noted a positive trend compared to baseline. In the last study it was reported that there was improvement in behavioural scores but the findings did not reach statistical significance (Cawley et al., 1994).

2.7. Rating of Studies

The ten studies were systematically assessed with the EPHPP tool (EPHPP, Quality Assessment Tool for Quantitative Studies, 2009) by both the first (SMA) and second author (CA) independently. Cohen's kappa was statistically significant (K = 0.40, p < 0.01) indicating at least moderate agreement between the reviewers. When there was a discrepancy between the reviewers' individual six component ratings (selection bias, study design, confounders, blinding, data collection method and withdrawals and dropouts), the reviewers examined the paper jointly and a consensus was agreed for both the individual six component ratings and the global rating. The individual and consensus ratings are recorded in Table 2 as well as reasons for discrepancies alongside any particular strengths of the individual study. Only two of the studies were given a global rating of moderate (i.e. only one weak component rating) for their quality (Heimlich, 2001; Cawley et al., 1994) and the remaining eight studies were rated as weak quality (i.e. two or more weak component ratings).

In summary when rating the six-part component ratings for individual studies most were found to be weak in quality when using the EPHPP tool: all of the studies were rated weak in identifying and controlling for confounders in the design or in the analysis. Nine studies were rated as weak in reporting numbers and reasons for drop outs/withdrawals from the study. Six studies were rated weak in data collection methods, namely the lack of valid and reliable measuring scales. Four studies were weak in blindness raters to outcomes (although, a significant proportion of the other studies did not report blinding status in their paper). Some of the papers that reported that assessors were blinded appeared to be referring to other raters in the study but not specify about being blind to participant intervention.

3. Discussion

3.1. Main Findings

There is international interest in this topic reflected in the included studies originating from various countries and common amongst them the purported positive psychosocial outcomes reported for people with ID with the addition of AAT in four areas: behaviour, social, cognitive and emotional. Social interaction/socialisation was the most frequently examined across all the studies accompanied with reports of statistical significance, as was emotional, behavioural and cognitive areas of outcomes.

Methodological Weaknesses. This review found that the majority of studies were of a weak design in examining the effectiveness of AAT for psychosocial outcomes for people with ID. Methodological problems included lack of controlling for confounders, non randomised designs, non reporting attrition and use of non standardised data outcome tools. The majority of studies did not report clearly about blinding of outcome raters, leaving studies open to potential observer bias. Furthermore, most of the studies had small sample sizes and did not report the severity of ID within their participant population. The quality assessment of the studies found that there were no strong quality studies, a fifth were of moderate quality with the rest being weak quality.

External Validity. Findings from the AAT studies in this review on psychosocial outcomes would be limited to the ID population, given that previous studies have noted differences of AAT effects between ‘disabled’ and ‘non-disabled’ groups (Nimer and Lundahl, 2007). The included studies were consistent in having an ID population (mainly children and adolescents) and the majority used similar study designs (pre and post intervention). However, the extent to which the generalisation of findings can be applied to the ID population as a whole is further restricted by a number of factors discussed below.

Small Number of Studies and Participant Numbers. There is a lack of studies in this population (especially adult population) and overall small participant numbers. We had ten studies included in our final review, which reflects on the small number of studies in the literature. This is consistent with findings of other recent reviews (Kamioki et al., 2014; Maujean et al., 2015) which included 11 and 7 papers respectively demonstrating the paucity of research in this area. A meta-analysis of effectiveness of AAT in depression had even less with just five included studies (Souter and Miller, 2007). The participant numbers in our included studies ranged from 3 to 23 with 30% having seven participants or less, which again is not unusual compared to other reviews in the field of AAT and targeted populations which have to draw upon a limited number of studies with sample sizes of: 6-64 with 27% having seven participants or less (Filan and Llewellyn-Jones, 2006), 5-25 with 20% having seven participants or less (Souter and Miller, 2007) and 1-42 with 36% having seven participants or less (O’Haire, 2013).

Male Gender Predominance, ID Severity and Co-morbid Conditions. The severity of ID was also not described in more than half of the studies which is important when comparing effects in the ID population, which already comprises a heterogeneous group. In particular challenging behaviour, notably aggression and self-injurious behaviour is more frequent in people with more severe ID (Xeniditis et al., 2001; Poppes et al., 2010). In the three studies in this review (Borioni et al., 2012; Memishevskij and Hodzhikj, 2010; Heimlich, 2001) which had recorded the severity of ID and examined behavioural
outcomes, the majority of participants (36 out of 41, 88%) had a level of ID that was moderate or below. Furthermore, there is evidence to suggest that the level of ID severity may influence outcome with AAT; one study (Chamrádová, 1995) who examined AAT in children with severe ID, reported that those children with a higher level of skills and mental age had better contact with the live dog and required less support from the therapist.

Most of the participants in the studies had higher numbers of male participants. This may be significant because one study (Memishevikj and Hodzhikj, 2010) found differences between females with mild ID and males with severe ID receiving AAT with the former showing improvement in speech, socialisation and sensory/cognitive functions. However due to small participant numbers the difference cannot be extended to the wider ID population. Further larger trials are required to determine if there are variations in outcome in AAT depending on gender or severity of ID. There is some evidence to suggest in other groups, such as the elderly, that pet interaction serves different roles for females and males (Miller et al., 1992).

Six of the studies also reported co-morbid conditions, mostly autism. This is important for behavioural interventions as outcomes in children with autism who also have co-morbid ID can differ to those without ID (Ben-Itzchak and Zachor, 2007). There is also evidence to suggest that children with autism that have higher functioning ability receiving intensive behavioural intervention have greater outcomes in intervention programs (Virués-Ortega et al., 2013).

**Inconsistency in Application of AAT Intervention.** The uniformity of AAT as an intervention varied considerably in duration, location, the administrator of therapy, type of animal and whether it was individually applied or in group settings. This is similar to the findings from the systematic review of AAT and ASD whereby the author adds that structured protocols/manuals may improve the consistency of AAT (O’Hare, 2013). Indeed, attempts to regulate AAT practice with training and qualifications exist through charities or organisation such as The European Society of Animal Assisted Therapy (http://www.en.esaat.org/). However, one downside to this is that by increasing regulatory requirements, requiring expense and time commitment, this may prevent the uptake of it or worse potentially stall research into the area. Perhaps more local, cheaper or sponsored alternatives are the best way to pursue this. A particular useful means of having health or social services sponsor training courses, qualifications and potentially services, would be by creating the effective evidence base for it. One way to help this is by conducting stronger quality methodological studies. In the UK, the charity Pets as Therapy (Patsastherapy.org) assists professionals undertaking research with animal intervention using PAT dogs and cats and they have been used in the National Health Service (NHS).

In addition, there may be a varying effect on psychosocial outcomes for people with ID depending on the type of animals used in AAT. One study in this review (Borioni et al., 2012) found differences between the groups receiving equestrian therapy (horses) and onotherapy (donkeys) with only the later group’s improvement in communication being statistically significant. Further larger trials are required to determine if there are differences in outcome in AAT depending on type of animal. Indeed, studies with general/mixed populations which have considered this question found effects sizes and confidence intervals more suggestive that dogs have the greatest effect (Nimer and Lundahl, 2007).

**Novelty and Hawthorne Effect.** Whether the animal represents a new or exciting event which is providing the desired outcome (a novelty effect) in AAT or whether it is the presence of the live animal itself, is something which has been considered a threat to construct validity of AAT studies (Marino, 2012). One of the studies in this review (Chamrádová, 1995) did attempt to control for novelty effects by using an artificial plastic dog compared to a real live dog and reported findings that the live dog group did better. This is in contrast to other studies in the field which have found no difference in outcome between groups with a robotic dog versus a live dog (Banks et al., 2008). Again this might be something which is distinct to the ID population and may be worth exploring with further studies.

Furthermore, whether it is the animal or the person administering the AAT that is a determining factor in outcomes is something which was not examined in any of the included studies but has been identified as a factor to be further investigated in the literature (Souter and Miller, 2007). Indeed, it can prove difficult to distil the precise active element which is effective in AAT. A recent systematic review recognising this, commented that AAT as an intervention is completely different to pharmacological or traditional rehabilitation methods and recommended that alternative interventions are added to the Consolidated Standards of Reporting Trials (CONSORT) 2010 checklist (Kamioki et al., 2014).

4. **Limitations**

This review has found consistent positive psychosocial outcomes for people with ID using AAT. However, this finding should be treated with caution and should not be seen as conclusive evidence of effectiveness due to the methodological diversity amongst studies in this population. Such heterogeneity in AAT literature reviews and meta-analyses is not uncommon, for example a recent review (Mauean et al., 2015) noted considerable variation in the measures used to assess psychological outcomes between the studies and in some reviews (Kamioki et al., 2014) it has prevented meta-analysis. In those studies that have completed meta-analyses of AAT (Nimer and Lundahl, 2007; Virués-Ortega et al., 2012), heterogeneity has been exacerbated by diverse target populations and the authors advised caution in interpreting the results. This review therefore attempted to have a more targeted population by applying strict criteria over ID diagnosis for included AAT studies, focusing exclusively on the ID population, something which had been lacking in the literature. However, there were many studies found through the literature search which fell short of having a substantial ID population, subgroup analysis for ID participants or did not report the number of ID participants.

Having strict inclusion criteria over ID diagnosis may have been a double edged sword; a strength which allowed better comparison between studies of the ID population reducing heterogeneity, yet a drawback as many studies with partial or
mixed populations consisting of some ID participants were excluded leaving a smaller base of studies from which to draw conclusions. As an additional layer of comprehensiveness, such papers which fell short of being included in the final stage due to ID diagnosis were acknowledged and tabulated in the appendix. Of such studies, two (Draper et al., 1990; Scholl et al., 2008) out of the nine studies contained adult only populations and another one study (Nathanson, 2007) contained a mixed children and young adult population.

Despite the stringency applied to studies over ID diagnosis, broader criteria were accepted in others areas such as age, participant number and study design. As a consequence this increased the heterogeneity of the study base placing limits on the extent to which the findings can be generalised amongst the ID population.

It is also important for further work in this area to note that ID terminology has changed over time and at different rates in various countries so that various search terms or keywords are not missed in literature searches. In this review, all of the American papers, up until 2008, used the term mental retardation. This has been regarded as an offensive word for many years and efforts have been made to reduce inappropriate usage of this term e.g. such as a website launched in 2008 by the Special Olympics (www.r-word.org), yet it still is the medical term which the World Health Organisation use and list in the ICD-10. Furthermore, studies may refer to individual syndromes which are always accompanied by a degree of ID. In this review we searched using multiple terms for ID as well as common syndromes. Learning Disabilities in USA papers will often equate to Learning Difficulties in the UK, but the term Learning Disabilities is still extensively used in the UK, was used by one of the included papers in this review (Limonid et al., 2007) and the term continues to be used by a leading charity for people with ID in the UK, Mencap, the Voice of Learning Disability (http://www.mencap.org.uk).

Terms such as AAT are neither uniformly defined in the literature for studies undertaken involving animals for therapeutic purpose, nor measures of its effectiveness. This is shown in the results by nearly all the studies using different names, with AAT being the exception and used the most times. This created difficulty in the search for studies that so many various terms had to be used to avoid missing significant studies. Despite this, the database searches identified just under half of the final included studies which may indicate the need for even more over-inclusive search terms. This is further substantiated by looking at the journals that the articles were published in Table 1 demonstrating the diversity in the subject, reflecting the spread of AAT as a topic across the literature and therefore requiring a comprehensive supplementation to the search beyond databases. This entails searching the grey literature and hand referencing, which is where the remaining studies were identified as potential studies for review. Whilst this was a strength of the study, an additional layer of comprehensiveness would have been to contact authors in the field. Another limitation of the review is a potential publication bias as only peer-reviewed published journal articles were included (excluding conference abstracts, doctoral theses etc).

One of the strengths of this review was to open the search to non-English published studies as both a meta-analysis of AAT (Nimer and Lundahl, 2007) and a systematic review of AAT in autism (O’Haire, 2013) restricted to English-only publications which may have created a language bias. In this review a non-English publication (Chamrádogá, 1995) was included amongst the final included studies. Using a standardised instrument such as the EPHPP Quality Assessment Tool to rate the quality of the studies was a particular strength of the review as well as using a second rater to improve the reliability of findings.

5. Further Work

From a general point of methodological rigor, what is required is to increase the quality of studies looking at AAT, ideally through multicentre randomised controlled trials allowing confounding factors to be eliminated at baseline. Such studies are more likely to detect a significant difference between groups and standardised data collection tools that are both valid and reliable to ensure consistency and accuracy in reporting outcomes. It is also equally important to report adverse events and withdrawals/drop outs as this may help future researchers with research design and or participants. Over all this will help reduce threats to internal validity and help with generalisation. Reporting of psychosocial outcomes can be complex and to reflect this there is work underway to extend the CONSORT to include social and psychological interventions (Mayo-Wilson et al., 2013).

More specifically, given that the one component of the psychosocial outcomes that was particularly lacking in studies was behavioural outcomes, it makes sense for more work to be done in this area, alongside examining whether type of animal, gender and severity of ID impact on outcomes with AAT in this population. Interventions that are evidence based may help ameliorate biases in health inequalities. For people with ID this may be particularly relevant as an estimated 10–20% of people with ID experience challenging behaviours (McClintock et al., 2003). The use of AAT in helping support existing therapies that work at managing challenging behaviours in people with ID e.g. psychological, behavioural, speech and language etc may be potentially beneficial and should be the focus of further research. Indeed, this is symptomatic of a larger problem within ID research that of few high quality trials being conducted and a call for evidence based research in this population to be a priority (Hastings, 2013). Beyond this, interventions should be examined not only for clinical effectiveness but also cost effectiveness, quality of life improvements as well as harms and benefits.

6. Conclusions

Current evidence shows that AAT may be a potentially useful supportive intervention in improving quality of life in persons with ID but good quality research is lacking. Therefore it merits further exploration within this population. Future
research should strive to address the methodological limitations of existing work if AAT is to be considered a clinical and cost effective adjunct in services for people with ID.

**Conflict of Interest**

None declared

**Acknowledgements**

We are grateful to: Mr Milan Klemš for translation assistance with the included Czech journal publication, Mrs Catherine Brown for translation assistance with retrieved French publications, Mrs Anna Bolinger-Maber for translation assistance with retrieved Polish publications and Dr Dawid Aleksandrowicz for translation assistance with retrieved German publications.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ridd.2015.12.005.

**Appendix C. Appendix**

**Table A1**

<table>
<thead>
<tr>
<th>Journal Reference</th>
<th>Therapy terminology (Animal intervention)</th>
<th>Diagnosis of group (intellectual disability participants in bold)</th>
<th>N (%) of intellectual disability participants</th>
<th>Age range of participants (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diits et al., 2011 <em>Journal of Creativity in Mental Health</em> USA</td>
<td>Dolphin-Assisted Therapy (Dolphin)</td>
<td><strong>Mental Retardation</strong>, Cerebral Palsy, Developmental Delay, Developmental Disorder, Developmental Deficits, West Syndrome, Autism, Hemiparesis, Encephalitis, de Morsier Syndrome, Down’s Syndrome, Epilepsy, Rett Syndrome, Microcephaly, Lesch-Nyhan Syndrome</td>
<td>Not reported</td>
<td>Age not reported</td>
</tr>
<tr>
<td>Breitenback et al., 2009 <em>Anthrozoo¨s</em> Germany</td>
<td>Dolphin-Assisted Therapy (Dolphin)</td>
<td>Down Syndrome, Autism, Mental and Physical Disabilities</td>
<td>15 (13%)</td>
<td>5-10</td>
</tr>
<tr>
<td>Scholl et al., 2008 <em>Therapeutic Communities</em> Austria</td>
<td>Animal-assisted therapy (Goats)</td>
<td>Deaf and Multiply Disabled, impaired vocal communication, mental impairments</td>
<td>Not reported</td>
<td>18-45</td>
</tr>
<tr>
<td>Nathanson, 2007 <em>Anthrozoo¨s</em> USA</td>
<td>Dolphin-Assisted Therapy (Dolphin)</td>
<td><strong>Mental Retardation</strong> due to anoxia or other causes, autism, cerebral palsy, Lennox Gastaut Syndrome, Microcephaly, Muscular Dystrophy, Pervasive Developmental Disorder, Rett Syndrome, Smith-Lemli-Opitz Syndrome, Tuberous Sclerosis</td>
<td>20 (57%)</td>
<td>3 – 20</td>
</tr>
<tr>
<td>Lukina, 1999 <em>Human Physiology</em> Ukraine</td>
<td>Dolphin-Assisted Therapy (Dolphin)</td>
<td>Mental Retardation, Autism</td>
<td>Not reported</td>
<td>4-10</td>
</tr>
<tr>
<td>Nathanson, 1998 <em>Anthrozoo¨s</em> USA</td>
<td>Dolphin-Assisted Therapy (Dolphin)</td>
<td>Down’s Syndrome, Angelman Syndrome, Tuberous Sclerosis, Microcephaly, Rett Syndrome, Cerebral Palsy, Hydrocephaly, Traumatic Brain Injury, Asperger’s Syndrome, Autism, Attention Deficit Disorder, Developmental Delay, Pervasive Developmental Delay, Rett Syndrome, Brain damage resulting in Mental Retardation</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Nathanson et al., 1997 <em>Anthrozoo¨s</em> USA</td>
<td>Dolphin-Assisted Therapy (Dolphin)</td>
<td>Rett Syndrome</td>
<td>23 (49%)</td>
<td>2-13</td>
</tr>
<tr>
<td>Draper et al., 1990 <em>Psychiatric Journal of the University of Ottawa Canada</em></td>
<td>Pet Facilitated Therapy (Dog)</td>
<td>Mental Retardation, Dementia and Post Traumatic Brain Syndrome</td>
<td>Not reported</td>
<td>20-87</td>
</tr>
<tr>
<td>Zemke et al., 1984 <em>Occupational Therapy in Mental Health</em> USA</td>
<td>Horseback Riding (Horse)</td>
<td>Learning Disabled, Educable Mental Retardation</td>
<td>Not reported</td>
<td>6-16</td>
</tr>
</tbody>
</table>
References


Further reading (Web References)


