Can findings from randomized controlled trials of social skills training in autism spectrum disorder be generalized? The neglected dimension of external validity

Autism
2016, Vol. 20(3) 295–305
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1362361315583817
aut.sagepub.com

Ulf Jonsson^{1,2,*}, Nora Choque Olsson^{1,3,*} and Sven Bölte^{1,3}

Abstract

Systematic reviews have traditionally focused on internal validity, while external validity often has been overlooked. In this study, we systematically reviewed determinants of external validity in the accumulated randomized controlled trials of social skills group interventions for children and adolescents with autism spectrum disorder. We extracted data clustered into six overarching themes: source population, included population, context, treatment provider, treatment intervention, and outcome. A total of 15 eligible randomized controlled trials were identified. The eligible population was typically limited to high-functioning school-aged children with autism spectrum disorder, and the included population was predominantly male and Caucasian. Scant information about the recruitment of participants was provided, and details about treatment providers and settings were sparse. It was not evident from the trials to what extent acquired social skills were enacted in everyday life and maintained over time. We conclude that the generalizability of the accumulated evidence is unclear and that the determinants of external validity are often inadequately reported. At this point, more effectiveness-oriented randomized controlled trials of equally high internal and external validity are needed. More attention to the determinants of external validity is warranted when this new generation of randomized controlled trials are planned and reported. We provide a tentative checklist for this purpose.

Keywords

child psychiatry, evidence-based, methods, neurodevelopmental disorders, systematic review

Introduction

Multiple strategies have been proposed to promote social communication skills in children and adolescents with autism spectrum disorder (ASD). Among these, the concept of social skills group interventions (SSGI) has attracted the most attention in the past 20 years (Reichow et al., 2012; Williams White et al., 2007). SSGI is an umbrella term comprising a wide range of teaching methods such as social stories, video modeling, social problem solving, pivotal response training, scripting procedures, computer-based teaching, prompting procedures, and self-monitoring (Bohlander et al., 2012). A recent systematic review of SSGI evidence identified five randomized controlled trials (RCTs) on school-aged children and adolescents with high-functioning ASD (Reichow et al., 2012). The review concluded that there is some evidence that SSGI can improve social skills in individuals with ASD.

Notably, the generalizability was deemed to be unclear, as all studies were conducted in the United States, focused mainly on children aged from 7 to 12 years, and included only participants of average to high intelligence (Reichow et al., 2012). Recent reviews focusing on methodological aspects of the accumulated research on SSGI raise

¹Department of Women's and Children's Health, Karolinska Institutet, Sweden

²The National Board of Health and Welfare, Sweden

³Center for Psychiatry Research, Stockholm County Council, Sweden

*Contributed equally

Corresponding author:

Ulf Jonsson, Center of Neurodevelopmental Disorders at Karolinska Institutet, Pediatric Neuropsychiatry Unit, Department of Women's and Children's Health, CAP Research Center, Karolinska Institutet, Gävlegatan 22 (Entré B), floor 8, SE-11330 Stockholm, Sweden.
Email: ulf.jonsson@ki.se

additional questions about generalizability, including how gains are generalized and maintained (Kaat and Lecavalier, 2014; McMahon et al., 2013). The methodological review by Kaat and Lecavalier also indicates that the determinants of external validity are often inadequately reported. For instance, they found that less than half of the identified studies had a well-characterized sample, defined as information about age and sex of participants and an independent evaluation of either intellectual functioning, or an independent confirmation of ASD diagnosis.

External validity refers to whether findings can be generalized across populations and environments. Reliable judgment about the generalizability of the accumulated body of RCTs is essential if interventions are to be successfully implemented in routine clinical practice. The external validity is sensitive to many factors, for instance, setting of the trial, selection of participants, characteristics of included participants, differences between the protocol and routine practice, choice of outcome measures, follow-up timing, providers' characteristics and training, treatment fidelity, number of participants, and trainer ratio (Rothwell, 2006).

Traditionally, the main target of systematic reviews is on the determinants of internal validity (i.e. the quality of design, conduct, and analysis), to ensure that the results are valid for the study population and setting (Higgins and Green, 2011). External validity, however, often receives insufficient attention in primary research studies and systematic reviews (Rothwell, 2006). The need for increased attention to such issues has been explicitly highlighted for psychosocial interventions (Gardner et al., 2013; Grant et al., 2013).

In this systematic review, we expand on previous reviews by fine-mapping the determinants of external validity of RCTs on SSGI reported in primary research studies, with the purpose to bring attention to these aspects when the next generation of RCTs are planned, evaluated, and reported in the scientific literature. We do this at a point in time when the SSGI research is entering a new phase, where the key question is how the results from early efficacy oriented trials are transported into real-world clinical settings. We evaluate different aspects (source population, included population, context, treatment provider, treatment intervention, and outcome) of the clinical generalizability for individual trials and the SSGI evidence base as a whole.

Material and methods

Protocol and registration

This study was based on a systematic review conducted by the authors of this article on behalf of the Swedish Board of Health and Welfare, registered in PROSPERO (International prospective register of systematic reviews;

crd.york.ac.uk/NIHR_PROSPERO/), registration no. CRD42013003780. In the commissioned report, four groups of interventions of relevance for children and adolescents with ASD were assessed: SSGI, early intensive behavioral intervention (EIBI), treatment and education of autistic and related communication handicapped children (TEACCH), and interventions involving relatives of children and adolescents with ASD. A combined search was used to address all four questions. In the present review, only the external validity pertaining to SSGI was examined.

Eligibility criteria

Participants. Children and adolescents with a diagnosis of ASD, autism, Asperger syndrome (AS), or pervasive developmental disorders-not otherwise specified (PDD-NOS) (as diagnosed using the DSM-IV, DSM-IV-TR, or ICD-10 criteria). Studies with adult populations (>18 years) were excluded.

Interventions. SSGI, defined as an intervention aimed at improving social skills characterized by participation of at least two individuals with ASD in therapy sessions led by one or more therapists.

Comparator. Any comparator.

Study design. Randomized controlled trial.

Information sources

Electronic searches were conducted using medical subject headings (MeSH) and relevant text word terms. Five databases (Medline, PubMed, PsycINFO, CINAHL, and ERIC) were searched, up to 19 December 2014.

Search strategy

The search was performed by a specialist at the Karolinska Institutet library. We used search terms relevant for the eligible interventions and the population. Search results were limited to original studies from 1990 or later, written in English, Danish, Norwegian, or Swedish. Animal studies and case studies were excluded. The updated search included only the search terms related to SSGI. For a detailed description of search terms related to SSGI, see Appendix 1, available online.

Study selection

Two reviewers independently screened the titles and abstracts of all the citations identified by the search strategy. Studies of potential relevance for SSGI were screened a second time by two reviewers. If deemed necessary at

this stage, the article was obtained in full text and two reviewers independently assessed them for inclusion. Any disagreements were resolved by discussions. Reference lists and systematic reviews were screened for additional studies of relevance. The citations identified in the updated search were screened in the same manner, with the exception that the search tool in EndNote was used first in order to identify citations including in any field either the truncated term random or the term RCT.

Data collection process

From each included study, data were extracted and inserted in a table by one of the authors. A second author checked the integrity of the data. Any disagreements were resolved by discussion. If no agreement could be reached, a third author would decide. In case the information was reported or partly reported, the relevant available information was inserted in the table. If the information was missing in the report, the cell was left empty.

It was also planned that study protocols referred to in the original publications would be searched for additional information. No other sources were searched for information.

Data items

In order to systematically evaluate external validity, a list of potential determinants was compiled. Items of relevance for external validity were selected from a review of reporting guidelines by Grant et al. (2013) and completed by additional items of clinical significance derived by face validity. Included items were clustered into six overarching themes related to external validity: source population, included population, context, treatment provider, treatment intervention, and outcome (Table 1).

Planned methods of analysis

For each coded item, the coding was summarized across all included studies. In a narrative synthesis, we summarized the extracted information for the whole body of included studies.

Results

In the first screening of abstracts, 217 unique citations were identified as of potential relevance for social skills training. A total of 199 of these reports were excluded: 57 case studies, 39 non-randomized studies, 75 studies with irrelevant research question, 13 studies of another intervention, 3 studies with adult participants, 6 study protocols, and 6 reviews (see Figure 1). In total, 15 eligible RCTs (reported in 18 publications) were identified (Baghdadli et al., 2013; Beaumont and Sofronoff, 2008;

Begeer et al., 2011; DeRosier et al., 2011; Frankel et al., 2010; Koenig et al., 2010; Koning et al., 2013; Laugeson et al., 2009; Lerner and Mikami, 2012; Lopata et al., 2010; Schohl et al., 2014; Solomon et al., 2004; Thomeer et al., 2012; White et al., 2013; Yoo et al., 2014). A long-term follow-up of the study by Frankel and colleagues was presented in a separate publication (Mandelberg et al., 2014). The study by Schohl et al. (2014) also included measures of parent and family outcome and electroencephalography (EEG), which were reported in separated publications (Karst et al., 2015; Van Hecke et al., 2015).

Source population

An ASD diagnosis ((AS), high-functioning Autism (HFA), or (PDD-NOS)) was a criterion for eligibility in all included trials. In most trials, the diagnosis was confirmed by standardized diagnostic tools or rating scales, such as the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 1999), Autism Diagnostic Interview-Revised (ADI-R) (Rutter et al., 2003), and Social Responsiveness Scale (SRS) (Bölte et al., 2008), but in two of the trials this was not clearly stated (Laugeson et al., 2009; Lopata et al., 2010). A minimum IQ or verbal IQ level (>60 –85) was an additional criterion for inclusion in all trials except one, in which only participants with HFA were eligible (Lerner and Mikami, 2012). Two trials also used a minimum score on the Comprehensive Assessment of Spoken Language as a criterion for inclusion (Lopata et al., 2010; Thomeer et al., 2012). In one trial (White et al., 2013), only individuals with a comorbid anxiety disorder were included.

All trials except one (Thomeer et al., 2012) provided some information about the recruitment procedure, although this information was insufficient. Two of the trials used public announcements (Beaumont and Sofronoff, 2008; Lopata et al., 2010) whereas the remaining 12 recruited from academic centers, clinics, schools, or local organizations. (For details, see Appendix 2(a), available online.)

Included population

Of the 15 trials, 9 reported the number of screened and randomized participants, as well as the number of participants who completed the study (Baghdadli et al., 2013; Begeer et al., 2011; Koenig et al., 2010; Laugeson et al., 2009; Lopata et al., 2010; Schohl et al., 2014; Thomeer et al., 2012; White et al., 2013; Yoo et al., 2014), while some of this information was omitted in the remaining trials. No trial provided clear information about the number of invited individuals. No information about the participants' expectations and preferences was included in any of the trials.

Table 1. Checklist of information to be included in trial reports to facilitate assessment of external validity, with information provided in one of the included trial reports given as an example.

Items	Comment	Example (information included in the trial report by Frankel et al., 2010) ^a
Source population		
1. Primary diagnosis	Criteria and methods used to confirm the diagnosis	Autism spectrum disorder (ASD) according to ADOS, ADI-R, and ASSQ
2. Additional inclusion criteria	Any additional inclusion criteria	Verbal IQ >60; able to switch topics in a conversation; adequate knowledge of rules in playing board games and common school yard games; attending second through fifth grade regular classroom
3. Exclusion criteria	Any exclusion criteria	Proscribed psychotropic medication; clinical seizure disorder; gross neurologic disease, or other medical disorder; thought disorder
4. Recruitment plan and type and procedure of recruitment (e.g. advertisement and clinic referral)	Detailed description of the recruitment plan enabling identification of the source population (e.g. the clinic's catchment area and the type of patients referred to the clinic)	Parents calling to the University of California, Los Angeles (UCLA) Outpatient Clinic about the social skills program or parents evaluated by the Assessment Core of the UCLA Center for Autism Research and Treatment
Included population		
1. Sample selection	The number of individuals approached for participation, screened, randomized, and completing the trial	Approached: not reported Screened: not reported Randomized: 76 Completed: 68
2. Age	Mean (SD) and range	6–11 years Treatment/control mean (SD)=8.6 (1.3)/8.5 (1.3)
3. Sex	Number of males and females included	58/10
4. Comorbidity	The psychiatric and somatic comorbidity of the participants	Not reported
5. Additional sample characteristics	Baseline characteristics of the sample, such as diagnosis, social skills, socioeconomic status, IQ, ethnicity, and ongoing treatments	ASD diagnosis: – Social skills: SSRS—assertion, treatment/control, mean (SD)=9.5 (2.8)/9.4 (3.4) IQ: verbal IQ, treatment/control, mean (SD)=106.9 (19.1)/100.5 (15.7) Parent education: algorithm combining level of education with occupation Ethnicity: Caucasian, treatment/control=77.1/54.5% Psychotropic medication: excluded
6. Treatment preferences and expectancies	Information about the participants treatment preferences and expectancies (e.g. choice of treatment and motivation)	Not reported
Context		
1. Location	The geographical area in which the study occurred (e.g. city, country, and region)	Los Angeles, CA, USA
2. Concurrent secular events/timing	The time period of data collection (e.g. month and year) and external events occurring at the time of the intervention that could influence outcome	Recruited from September 2003 to March 2008
3. Setting type/service environment	The type of settings of data collection in the study (e.g. classroom, participants' homes, clinic, university clinic); service provider, characteristics of the service environment (e.g. economic, legal, political, demographic, technological, and policy-related environment; availability of alternatives outside the trial context; compensation structures; unique features of the trial environment)	Regular outpatient clinical service
4. Number of settings	Different settings of data collection in the study (e.g. number of classrooms, participants' homes, and clinics)	Not reported
5. Ethics	Information, consent	Approved by the university and the NIMH IRB. Parent consent and child assent
6. Incentives	Any incentives or compensation for participating in the trial	Not reported

Table 1. (Continued)

Items	Comment	Example (information included in the trial report by Frankel et al., 2010) ^a
Treatment provider		
1. Number of providers	Number of professionals and teams delivering the intervention	Teams: 1 (+1 for parent groups) Therapists/team: 1 group leader
2. Staffing	Actual staffing of the intervention (i.e. number and qualification of the staff involved)	PhD level psychologist with >10 years' experience (licensed clinical social worker with >5 years' experience in the parent group)
3. Provider training	How intervention providers were trained	10 years of experience in social skills training
4. Supervision	Any contacts between intervention providers and supervisors/researchers	Not reported
5. Treatment fidelity	Steps to measure adherence of care providers with the protocol (e.g. incentives for staff compliance, participant feedback, mailings, or phone reminders)	Fidelity checklists covering the primary content of the protocol were created for each treatment session
6. Provider preferences	Information about treatment preferences of providers (if more than one treatment option is available)	Not reported
Treatment intervention		
1. Intervention manual	Information for accessing intervention materials (e.g. protocol or manual) to allow for replication	Parent-assisted Children's Friendship Training (CFT). Parents are integrated within separate concurrent sessions
2. Composition of groups	Number of groups and group size	Usually 10 in each group, whereof a maximum of four with ASD
3. Duration	The intended length of the intervention (e.g. 20-week program)	12 weeks
4. Frequency	The frequency of intervention (e.g. one per week)	Weekly
5. Intensity	The intended intensity of the intervention (i.e. length of each session)	60 min
6. Costs	Any costs associated with the treatment (e.g. cost related to material, staff, supervision, and training)	Not reported
7. Deviation/tailoring	Intervention adaptation by researchers and staff (e.g. types and extent of deviations from protocol that have not been highlighted above, including supplementing the treatment)	Verification that no other topic was presented other than those in the manual
Outcome		
1. Type of data	Qualitative and/or quantitative, scales, tests, and observations	Quantitative data Rating scales
2. Informants and measures	The primary and secondary outcome measures (including informant for each measure) and the informant for each measure (e.g. child, parent, staff, and teacher)	Child: The Loneliness Scale, Piers-Harris Self-Concept Scale Parent: QPQ, SSRS (primary outcome) Teacher: PEI
3. Generalizability and quality of outcome measures	Method of data collection, on-site results for reliability and validity, enactment of learned skills in relevant real-life settings	Multiple sources: yes Observations by blinded raters: no Any blinded assessment: no On site reliability or validity data: Yes, for QPQ Enactment: –
4. Adverse effects	Any harmful or unwanted effects (e.g. depressive symptoms, conflicts, sense of failure, stress, or adverse events)	Not reported
5. Timing of measurement	The timing of measurement and follow-up period for all groups and measures in the trial	Just prior to receiving the intervention (T1), the last night of the intervention (T2), and 12 weeks after the conclusion of the intervention (T3). Follow-up after 1–5 years presented in a more recent publication
6. Authors' view on generalizability	Any discussion referring to generalizability of the results	Findings limited to a select subsample of high-functioning children with ASD

^aThe largest included trial was selected as an example. Detailed information about all included trials in Appendix 2 (available online).

ADI-R: The Autism Diagnostic Interview-Revised; ADOS: The Autism Diagnostic Observation Schedule; ASSQ: Autism Spectrum Screening Questionnaire; NIMH IRB: National Institute of Mental Health Institutional Review Board; PEI: The Pupil Evaluation Inventory; QPQ: The Quality of Play Questionnaire; SD: standard deviation; SSRS: social skills rating system.

Figure 1. PRISMA flow-chart.

In total, 11 trials included children within an age range spanning from 6 to 13 years, while 4 trials focused mainly on adolescents (Laugeson et al., 2009; Schohl et al., 2014; White et al., 2013; Yoo et al., 2014). Only a minority of females was included in the trials and no study presented gender-specific results. Two reports included some information about psychiatric comorbidity among the included participants (Begeer et al., 2011; White et al., 2013), while the rest did not.

The sample characteristics presented varied across studies, but information about the participants' diagnosis, ethnicity, IQ, and parental education was frequently reported. Some trials mainly included participants with a diagnosis of PDD-NOS (Begeer et al., 2011; Koenig et al., 2010), while HFA and AS were the most common diagnoses in other trials. A number of different measures of social skills were used in the trials, complicating assessment of differences in severity of the target symptom between studies. The available information suggests that the participants were predominantly Caucasian and had parents

with relatively high educational level. (For details, see Appendix 2(b), available online.)

Context

One trial was conducted in Australia (Beaumont and Sofronoff, 2008), one in Canada (Koning et al., 2013), one in France (Baghdadli et al., 2013), one in the Netherlands (Begeer et al., 2011), one in South Korea (Yoo et al., 2014), and the remaining in the United States. The setting was described briefly or not at all: four trials were set at universities (Beaumont and Sofronoff, 2008; Lopata et al., 2010; Solomon et al., 2004; Thomeer et al., 2012), five in clinics, research clinics, or school (Baghdadli et al., 2013; Begeer et al., 2011; Frankel et al., 2010; Koenig et al., 2010; Schohl et al., 2014), one within family service (DeRosier et al., 2011), and five reports did not provide explicit information about setting (Koning et al., 2013; Laugeson et al., 2009; Lerner and Mikami, 2012; White et al., 2013; Yoo et al., 2014).

Incentives included a light meal and paid parking (Laugeson et al., 2009), a honorarium of US\$50.00 (Koning et al., 2013), and no fee for participation (Koenig et al., 2010). External events occurring at the time when the trials were conducted were not mentioned in any of the reports. (For details, see Appendix 2(c), available online.)

Treatment provider

Generally, information about staffing was scarce. The number of trainers in each skill-training group ranged from one to four across studies. The qualifications of the trainers varied: in several studies post-graduate, graduate, or undergraduate students delivered the intervention (Beaumont and Sofronoff, 2008; Laugeson et al., 2009; Lopata et al., 2010; Schohl et al., 2014; Thomeer et al., 2012; White et al., 2013; Yoo et al., 2014), while other studies also used clinical psychologists, psychotherapists, psychiatrists, social workers, registered nurses, and speech and language pathologists. The experience of the trainers varied from considerable practical training in conducting social skills groups in ASD to less experience. In total, 10 trials reported that providers received specific training in delivering the intervention (Baghdadli et al., 2013; Begeer et al., 2011; DeRosier et al., 2011; Laugeson et al., 2009; Lerner and Mikami, 2012; Lopata et al., 2010; Schohl et al., 2014; Thomeer et al., 2012; White et al., 2013; Yoo et al., 2014), while 5 trials did not include such information. Eight trials reported that the trainers received supervision (Baghdadli et al., 2013; Begeer et al., 2011; Laugeson et al., 2009; Lerner and Mikami, 2012; Schohl et al., 2014; Thomeer et al., 2012; White et al., 2013; Yoo et al., 2014). Treatment fidelity and adherence to the treatment protocol was controlled in some way in all trials. (For details, see Appendix 2(d), available online.)

Treatment intervention

All reports explicitly referred to a treatment manual or gave relatively detailed descriptions of the interventions. The treatments varied substantially across studies and focused on several aspects of social skills, including verbal communication (e.g. tone of voice and nonliteral language), non-verbal communication (e.g. eye contact, facial expressions, posture, and gestures), and social interaction (e.g. initiation, maintaining or leaving a social interaction, conversation skills and social problem solving, conflict in relationships, coping with bullying and teasing, self-regulation, and self-monitoring controlling negative emotions). However, only a few trials gave a detailed description of the number of groups treated and the number of participants in each group (Lopata et al., 2010; Solomon et al., 2004; Thomeer et al., 2012; Yoo et al., 2014).

The duration of the interventions varied between 4 and 20 weeks. Two trials had less than 10 sessions (Beaumont

and Sofronoff, 2008; Lerner and Mikami, 2012), two trials used a high intensity training with five daily treatment cycles 5 days a week for 5 weeks (Lopata et al., 2010; Thomeer et al., 2012), and the remaining trials ranged between 12 and 20 sessions. The intensity of the sessions ranged from 60 to 75 min to 2 h. (For details, see Appendix 2(e), available online.)

Outcome

A range of different measures of social skills was used across the studies, including rating scales, tests, and observations. All trials except one (Koenig et al., 2010) used more than one source of information (e.g. the child, parents, teachers, staff, or tests). Two trials used blinded observations (Koning et al., 2013; Lerner and Mikami, 2012), while an additional five trials included any form of blinded assessment (Baghdadli et al., 2013; Koenig et al., 2010; Laugeson et al., 2009; Schohl et al., 2014; White et al., 2013). Five trials reported follow-up assessments after 2–5 months (Beaumont and Sofronoff, 2008; Frankel et al., 2010; Laugeson et al., 2009; Thomeer et al., 2012; Yoo et al., 2014), while the other trials only included pre- and post-measurements.

A few studies used tools to measure symptoms of depression or anxiety (Schohl et al., 2014; Solomon et al., 2004; White et al., 2013; Yoo et al., 2014). One trial included global assessment of everyday functioning (White et al., 2013) and two trials reported change in clinical global impression (Koenig et al., 2010; White et al., 2013).

In the discussion sections of all studies, the authors refer to limitations related to external validity, such as restricted samples, no follow-up assessments, no evaluation of the transportability to other settings, and no information about generalizability of gains. For details, see Appendix 2(f), available online.

Discussion

For many years, ASD intervention research was criticized for a lack of evidence-based practice. Recently, several systematic reviews of SSGI in ASD identified a range of RCTs showing reasonable internal validity. However, the clinical applicability and implementability of these interventions has not been rigorously investigated yet. We examined information provided in RCTs of SSGI in ASD to judge the external validity of these trials. This information is pivotal to address questions raised by clinicians, clients, parents, policy makers, and other stakeholders related to implementation: To whom do the results apply? To what kind of settings do the results apply? Who can provide the treatment and how should the treatment be delivered? Are the gains generalized to the child's everyday environment and maintained over time?

Our review suggests that the generalizability of the evaluation research within this field is unclear. Moreover, information necessary for proper assessment of generalizability is often missing in the reports. Notably, the scant information about the pathway to recruitment precluded identification of the source population from which the participants were drawn. Unless detailed information about the source population is available (e.g. the clinic's patient mix and catchment area), general conclusions about generalizability are undermined.

A homogeneous population allows for a higher internal validity, but highly selective eligibility criteria can also considerably reduce the applicability of the trial results (Khorsan and Crawford, 2014; Rothwell, 2006). The eligibility criteria reported in the included trials mainly concerned age, IQ, aggressive behavior, and severe mental health problems. These eligibility criteria could be justified, provided that the intervention primarily is intended for high-functioning school-aged children and adolescents. Nevertheless, the study populations might still be highly selective, not representing clinical reality.

Detailed reporting of baseline characteristics of participants included in RCTs should allow clinicians to compare similarity with their actual patients in daily routine. Although baseline characteristics were described in almost all reports, the majority of trials did not take into account the high rates of comorbid conditions among persons with ASD, which is an important aspect when assessing to whom the results can be applied (Fortin et al., 2006; Matson and Cervantes, 2014; Rothwell, 2006). ASD as a heterogeneous condition with a spectrum of symptom severity and individual differences can have significant impact on the treatment effect (Bölte, 2014). Notably, some trials mainly included participants with a diagnosis of PDD-NOS. The use of PDD-NOS as a diagnostic category may vary between countries depending on diagnostic routines, traditions, and applied diagnostic systems (Elsabbagh et al., 2012; Huerta et al., 2012).

Some of the sample characteristics suggest that the samples were highly selective ASD individuals: For instance, most parents were relatively highly educated according to the available information. It is unclear to what extent the included participants were those with highly motivated and dedicated parents, as this is hardly ever assessed. Overall, there is a need for studies examining whether participant and intervention characteristics mediate/moderate treatment efficacy, as highlighted by a recent review (McMahon et al., 2013).

The feasibility and effects of some psychological interventions could depend on cultural aspects and it might be necessary to adapt the interventions according to cultural differences before implementation. Unfortunately, psychological interventions in general are rarely evaluated outside North America and Europe (Arnberg et al., 2013). Notably, one of the included trials was set in South Korea

(Yoo et al., 2014) and indicated that the intervention with modest cultural adjustments was efficacious in this cultural setting.

With respect to social skills outcome measures, different tools related to different component of social skills were used. Since "social skills" are described as a complex and multidimensional construct, it can be difficult to adequately assess improvements in social skills (Koenig et al., 2009; McMahon et al., 2013; Williams White et al., 2007). Doing so may require a multi-method measurement approach. Another issue is that few studies included blinded observations in the participants' everyday environment. Also, follow-up of long-term effects were not regularly reported. Only one trial report stated clearly that spontaneously reported adverse events were logged and reviewed (White et al., 2013). Missing information about potential adverse effects is a general shortcoming in RCTs of psychological interventions (Jonsson et al., 2014).

Treatment preferences might have an effect on the outcome (Swift and Callahan, 2009), but no report included information about the participants' or providers' preferences. Such information can be particularly important when two or more active treatments are compared. In addition, no information about costs related to the intervention was provided. This is unfortunate, since economic aspects of interventions are of paramount importance for implementation likelihood (National Institute for Health and Care Excellence, 2013).

Vital information was missing from all the reviewed research reports. This is understandable, given all the information that is demanded and the limited space provided by many journals. However, inadequate information about the determinants of external validity could considerably reduce the applicability of the trial results and ultimately lead to a waste of resources allocated to research (Glasziou et al., 2014). There are signs of improvement, though. Several recent research protocols for RCTs of SSGI suggest an increased focus on external validity (Choque Olsson and Bölte, 2013; Dekker et al., 2014; Freitag et al., 2013). In order to facilitate further improvement and a next generation of RCTs of SSGI in ASD accounting equally for internal and external validity, we suggest the list of items we used for this review as a tentative checklist (Table 1). Previous efforts to improve the reporting of results of RCTs, such as the CONSORT recommendations (Schulz et al., 2010), have mainly focused on internal validity. This checklist focuses exclusively on the determinants of external validity. We believe that the checklist can be useful for both researchers and reviewers.

It should be emphasized that the setting, care providers, and centers have obvious implications for external validity (Rothwell, 2006; Slack and Draugalis, 2001). In order to increase ecological validity, it is important that applied research uses a bottom-up approach with the intervention embedded in real clinical setting and

delivered by clinicians, teachers, or other practitioners who have direct contact with individuals with ASD in their everyday environment (Williams White et al., 2007). Top-down intervention research in academic setting reflecting no or limited clinical reality may have little impact on clinical practice, as it does not face and comply with the challenges of true clinical service.

Ensuring generalizability to the everyday clinical settings might compromise the internal validity. This “trade-off” is reflected in the differentiation between efficacy (how the intervention performs under ideal and controlled conditions) and effectiveness (how the intervention performs in real-world settings). A primacy of internal validity and efficacy is often appropriate in the first place for two reasons. First, initial trials might be of low scientific rigor and therefore have limited internal validity. Second, internal validity is an indispensable prerequisite for external validity (not vice versa). Once an intervention is proven efficacious, a natural next step is to continue with effectiveness studies, combining satisfactory internal validity with a high degree of external validity. As the quantity of SSGI RCTs is increasing continuously, a gradual shift or enlargement of the focus into effectiveness research seems now warranted in order to stimulate a new generation of RCTs of equally high internal and external validity.

A possible limitation of the present review is that it was restricted to RCTs. It is important to note that other designs than RCTs can give valuable insight into generalizability (Wells, 1999). Still, the results must have a satisfactory internal validity and sufficient information must be available regarding the setting, participants, and intervention. Many valuable one-group and quasi-experimental SSGI trials have been published in ASD and paved the way for RCTs (Kaat and Lecavalier, 2014; McMahon et al., 2013). These studies may appear more clinically naturalistic than RCTs. However, it is technically impossible that studies with poor internal validity have high external validity. It is evident from the review by Kaat and Lecavalier (2014) that extant non-randomized studies either had a small sample size or did not have a well-characterized sample, suggesting that they would add little to our understanding of generalizability.

An additional limitation is that we consistent with standard systematic review procedures only reviewed information provided in the primary research report. We did not contact the authors for missing information, nor did we review all possibly available literature and information outside of scientific journals. In particular, more detailed information about the intervention could be available in manuals or educational material. However, we believe that the information we checked for should be provided in the primary research reports. We also did not investigate a number of other factors necessary for successful implementation, including availability of the manuals, educational material, and training courses.

Conclusion

Trial reports should allow clinicians to judge to whom and in which context these results could reasonably be applied and how the treatment should be provided in order to obtain similar effects. In order to facilitate this, journal editors, reviewers, and researchers within this research field should pay more attention to information of relevance for external validity. In addition, more effectiveness-oriented RCTs of equally high internal and external validity are needed. This will provide a better evidence base for decisions about implementation into regular care.

Acknowledgements

The authors would like to thank Viviann Nordin for her contribution to the selection of studies in the systematic literature review, as well as Carl Gornitzki and Klas Moberg for their assistance with the literature search.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This study was supported by the National Board of Health and Welfare (Socialstyrelsen), Barnforskningen Astrid Lindgrens Barnsjukhus, Stiftelsen Sunnerdahls Handikappfond, Majblomman, Sällskapet Barnavård, and the Stockholm County Council. Sven Bölte was supported by the Swedish Research Council (grant no. 523-2009-7054).

References

- Arnberg FK, Alaie I, Parling T, et al. (2013) Recent randomized controlled trials of psychological interventions in health-care: a review of their quantity, scope, and characteristics. *Journal of Psychosomatic Research* 75(5): 401–408.
- Baghdadli A, Brisot J, Henry V, et al. (2013) Social skills improvement in children with high-functioning autism: a pilot randomized controlled trial. *European Child & Adolescent Psychiatry* 22(7): 433–442.
- Beaumont R and Sofronoff K (2008) A multi-component social skills intervention for children with Asperger syndrome: the Junior Detective Training Program. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 49(7): 743–753.
- Begeer S, Gevers C, Clifford P, et al. (2011) Theory of Mind training in children with autism: a randomized controlled trial. *Journal of Autism and Developmental Disorders* 41(8): 997–1006.
- Bohlander AJ, Orlich F and Varley CK (2012) Social skills training for children with autism. *Pediatric Clinics of North America* 59(1): 165–174, xii.
- Bölte S (2014) Is autism curable? *Developmental Medicine and Child Neurology* 56(10): 927–931.
- Bölte S, Poustka F and Constantino JN (2008) Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). *Autism Research* 1(6): 354–363.

- Choque Olsson N and Bölte S (2013) RCT multicenter study on social skills group training for children and adolescents with high-functioning autism spectrum disorder. Available at: <http://clinicaltrials.gov/show/NCT01854346>
- Dekker V, Nauta MH, Mulder EJ, et al. (2014) A randomized controlled study of a social skills training for preadolescent children with autism spectrum disorders: generalization of skills by training parents and teachers? *BMC Psychiatry* 14: 189.
- DeRosier ME, Swick DC, Davis NO, et al. (2011) The efficacy of a social skills group intervention for improving social behaviors in children with high functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders* 41(8): 1033–1043.
- Elsabbagh M, Divan G, Koh YJ, et al. (2012) Global prevalence of autism and other pervasive developmental disorders. *Autism Research* 5(3): 160–179.
- Fortin M, Dionne J, Pinho G, et al. (2006) Randomized controlled trials: do they have external validity for patients with multiple comorbidities? *Annals of Family Medicine* 4(2): 104–108.
- Frankel F, Myatt R, Sugar C, et al. (2010) A randomized controlled study of parent-assisted children's friendship training with children having autism spectrum disorders. *Journal of Autism and Developmental Disorders* 40(7): 827–842.
- Freitag CM, Cholemkery H, Elsuni L, et al. (2013) The group-based social skills training SOSTA-FRA in children and adolescents with high functioning autism spectrum disorder—study protocol of the randomised, multi-centre controlled SOSTA—net trial. *Trials* 14: 6.
- Gardner F, Mayo-Wilson E, Montgomery P, et al. (2013) The need for new guidelines to improve the reporting of trials in child and adolescent mental health. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 54(7): 810–812.
- Glaziou P, Altman DG, Bossuyt P, et al. (2014) Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet* 383(9913): 267–276.
- Grant SP, Mayo-Wilson E, Melendez-Torres GJ, et al. (2013) Reporting quality of social and psychological intervention trials: a systematic review of reporting guidelines and trial publications. *PLoS ONE* 8(5): e65442.
- Higgins JPT and Green S (2011) *Cochrane Handbook for Systematic Reviews of Interventions* (version 5.1.0). Available at: <http://handbook.cochrane.org/>
- Huerta M, Bishop SL, Duncan A, et al. (2012) Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *American Journal of Psychiatry* 169(10): 1056–1064.
- Jonsson U, Alaie I, Parling T, et al. (2014) Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: a review of current practice. *Contemporary Clinical Trials* 38(1): 1–8.
- Kaat AJ and Lecavalier L (2014) Group-based social skills treatment: a methodological review. *Research in Autism Spectrum Disorders* 8: 14–24.
- Karst JS, Van Hecke AV, Carson AM, et al. (2015) Parent and family outcomes of PEERS: a social skills intervention for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 45: 752–765.
- Khorsan R and Crawford C (2014) How to assess the external validity and model validity of therapeutic trials: a conceptual approach to systematic review methodology. *Evidence-Based Complementary and Alternative Medicine* 2014: 694804.
- Koenig K, De Los Reyes A, Cicchetti D, et al. (2009) Group intervention to promote social skills in school-age children with pervasive developmental disorders: reconsidering efficacy. *Journal of Autism and Developmental Disorders* 39(8): 1163–1172.
- Koenig K, White SW, Pachler M, et al. (2010) Promoting social skill development in children with pervasive developmental disorders: a feasibility and efficacy study. *Journal of Autism and Developmental Disorders* 40(10): 1209–1218.
- Koning C, Magill-Evans J, Volden J, et al. (2013) Efficacy of cognitive behavior therapy-based social skills intervention for school-aged boys with autism spectrum disorders. *Research in Autism Spectrum Disorders* 7: 1282–1290.
- Laugeson EA, Frankel F, Mogil C, et al. (2009) Parent-assisted social skills training to improve friendships in teens with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 39(4): 596–606.
- Lerner MD and Mikami AY (2012) A preliminary randomized controlled trial of two social skills interventions for youth with high-functioning autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* 27(3): 147–157.
- Lopata C, Thomeer ML, Volker MA, et al. (2010) RCT of a manualized social treatment for high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders* 40(11): 1297–1310.
- Lord C, Rutter M, DiLavore P, et al. (1999) *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Publishing.
- McMahon CM, Lerner MD and Britton N (2013) Group-based social skills interventions for adolescents with higher-functioning autism spectrum disorder: a review and looking to the future. *Adolescent Health, Medicine and Therapeutics* 2013(4): 23–28.
- Mandelberg J, Frankel F, Cunningham T, et al. (2014) Long-term outcomes of parent-assisted social skills intervention for high-functioning children with autism spectrum disorders. *Autism* 18(3): 255–263.
- Matson JL and Cervantes PE (2014) Commonly studied comorbid psychopathologies among persons with autism spectrum disorder. *Research in Developmental Disabilities* 35(5): 952–962.
- National Institute for Health and Care Excellence (2013) *Autism: the management and support of children and young people on the autism spectrum*. National Clinical Guideline 170. Available at: <http://www.nice.org.uk/guidance/cg170/resources/guidance-autism-pdf>
- Reichow B, Steiner AM and Volkmar F (2012) Social skills groups for people aged 6 to 21 with autism spectrum disorders (ASD). *The Cochrane Database of Systematic Reviews* 7: CD008511.
- Rothwell PM (2006) Factors that can affect the external validity of randomised controlled trials. *PLoS Clinical Trials* 1(1): e9.
- Rutter M, Le Couteur A and Lord C (2003) *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services.

- Schohl KA, Van Hecke AV, Carson AM, et al. (2014) A replication and extension of the PEERS intervention: examining effects on social skills and social anxiety in adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 44(3): 532–545.
- Schulz KF, Altman DG, Moher D, et al. (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine* 152(11): 726–732.
- Slack MK and Draugalis JR (2001) Establishing the internal and external validity of experimental studies. *American Journal of Health-System Pharmacy* 58(22): 2173–2181; quiz 2182–2173.
- Solomon M, Goodlin-Jones BL and Anders TF (2004) A social adjustment enhancement intervention for high functioning autism, Asperger's syndrome, and pervasive developmental disorder NOS. *Journal of Autism and Developmental Disorders* 34(6): 649–668.
- Swift JK and Callahan JL (2009) The impact of client treatment preferences on outcome: a meta-analysis. *Journal of Clinical Psychology* 65(4): 368–381.
- Thomeer ML, Lopata C, Volker MA, et al. (2012) Randomized clinical trial replication of a psychosocial treatment for children with high-functioning autism spectrum disorders. *Psychology in the Schools* 49(10): 942–954.
- Van Hecke AV, Stevens S, Carson AM, et al. (2015) Measuring the plasticity of social approach: a randomized controlled trial of the effects of the PEERS intervention on EEG asymmetry in adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 45(2): 316–335.
- Wells KB (1999) Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry* 156(1): 5–10.
- White SW, Ollendick T, Albano AM, et al. (2013) Randomized controlled trial: multimodal anxiety and social skill intervention for adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 43(2): 382–394.
- Williams White S, Keonig K and Scahill L (2007) Social skills development in children with autism spectrum disorders: a review of the intervention research. *Journal of Autism and Developmental Disorders* 37(10): 1858–1868.
- Yoo HJ, Bahn G, Cho IH, et al. (2014) A randomized controlled trial of the Korean version of the PEERS((R)) parent-assisted social skills training program for teens with ASD. *Autism Research* 7(1): 145–161.