The set-shifting profiles of anorexia nervosa and autism spectrum disorder using the Wisconsin Card Sorting Test: a systematic review and meta-analysis

H. Westwood1, D. Stahl2, W. Mandy3 and K. Tchanturia1,4*

1 Psychological Medicine, King’s College London, Institute of Psychiatry, Psychology & Neuroscience, London, UK
2 Department of Biostatistics, King’s College London, Institute of Psychiatry, Psychology & Neuroscience, London, UK
3 Ilia State University, Tbilisi, Georgia
4 University College London, Research Department of Clinical, Educational and Health Psychology, London, UK

Difficulties in set-shifting are commonly reported in both autism spectrum disorder (ASD) and anorexia nervosa (AN) populations. Despite this, it is not known whether this cognitive profile persists across different ages, or whether the profiles seen in ASD and AN are comparable. This systematic review and meta-analyses aimed to compare the set-shifting profiles, as measured by the Wisconsin Card Sorting Test (WCST) in adults and younger people with either ASD or AN, relative to healthy controls (HCs) and to statistically compare performance on the WCST between ASD and AN. In all, 24 studies on ASD and 22 studies on AN were identified. In ASD, there were significant differences between the clinical and HCs, with the ASD group making significantly more perseverative errors, indicating greater difficulty in set-shifting [pooled effect size of \( d = 0.67, 95\% \text{ confidence interval (CI)} \ 0.53–0.81, p \leq 0.001 \)]. This effect was consistent across the age span. For AN studies, there was a significant difference between adults with AN and HCs (\( d = 0.52, 95\% \text{ CI} 0.36–0.68, p \leq 0.001 \)) but a non-significant effect in child studies (\( d = 0.25, 95\% \text{ CI} 0.05 to 0.55, z = 1.66, p = 0.096 \)). Meta-regression indicated no effect of diagnosis (AN or ASD) on performance in adult studies but there was a non-significant trend (\( p = 0.053 \)) towards children with ASD performing worse than children with AN. While difficulties with set-shifting appear to be stable in ASD, there may be differences between children and adults with AN, which warrant further investigation.

Received 7 July 2015; Revised 9 February 2016; Accepted 16 February 2016; First published online 25 April 2016

Key words: Anorexia nervosa, autism spectrum disorder, meta-analyses, set-shifting, Wisconsin Card Sorting Test.

Introduction

A growing body of literature indicates a link between anorexia nervosa (AN) and autism spectrum disorder (ASD). A systematic review of the prevalence of ASD in AN reported a mean rate of 23% (Huke et al. 2013); however, six of the included studies came from the same Swedish community sample and this line of research is ongoing (Nielsen et al. 2015). There are also similarities in cognitive style between AN and ASD (Gillberg et al. 2010; Oldershaw et al. 2011; Courty et al. 2013; Tchanturia et al. 2013a) reported in the literature. Although elevated levels of autistic traits have consistently been found within AN populations (Baron-Cohen et al. 2013; Tchanturia et al. 2013b; Huke et al. 2014; Rhind et al. 2014; Westwood et al. 2016) it is not known whether the presence of these traits represents a shared, underlying mechanism (Zucker et al. 2007; Odent, 2010; Allely, 2013) or whether individuals with AN come to resemble those with ASD due to the chronic, starved state associated with the disorder (Pellicano & Hiller, 2013; Treasure, 2013). Direct comparison of the seemingly shared traits, as well as systematic exploration of these traits across stages of illness, e.g. childhood and adulthood, will be essential to understanding the extent to which these traits may be shared mechanisms in both AN and ASD.

Interest in the potential relationship between AN and ASD has led to examination of cognitive traits, believed to underpin ASD in AN. While some traits associated with ASD, such as emotional theory of mind, have been found to reduce in recovered AN groups (Oldershaw et al. 2010), other traits including difficulties with set-shifting, or flexible thinking, appear to persist after weight gain (Tchanturia et al. 2004; Gillberg et al. 2010; Danner et al. 2012; Lindner et al. 2014). Recent reports suggest that set-shifting inefficiencies are present in children with AN who have had a relatively short duration of illness (Lang...
et al. 2015a) and relatives with AN (Holliday et al. 2005; Roberts et al. 2010; Lang et al. 2015b). Conversely, fasting in healthy controls (HCs) has been found to reduce set-shifting test performance, suggesting that starvation may exacerbate underlying difficulties in this domain (Bolton et al. 2014; Pender et al. 2014). Thus, despite evidence of cognitive similarity between AN and ASD, the extent and nature of this likeness remain unclear. Exploring differences in the set-shifting profiles of children and adults with AN would speak to debates of whether neurocognitive traits such as set-shifting underlie the eating disorder or whether these cognitive traits are more related to eating disorder pathology such as starvation.

Differences in set-shifting have been observed between children and adults with AN. A systematic review of set-shifting ability, comparing adults with AN with HCs identified five papers employing the Wisconsin Card Sorting Test (WCST) (Heaton et al. 1993), a widely used measure of set-shifting, with a meta-analysis reporting a significant small effect size ($d = 0.36$; Roberts et al. 2007). A further meta-analysis of studies in children and adolescents with AN (Lang et al. 2014), however, found a non-significant pooled effect size ($d = 0.20$). A more recent study comparing children and adolescents with AN with HCs using a larger sample size (AN = 41, HC = 43) found that young people with AN made significantly more perseverative errors than HCs ($d = 0.49$; Lang et al. 2015a), suggesting that this cognitive profile is present early on in the disorder before the effects of starvation become prominent. A meta-analysis by Wu et al. (2014) included a range of neuropsychological tasks assessing set-shifting, rather than just the WCST. This revealed a medium effect size for inefficient set-shifting in restrictive AN ($g = -0.51$) but no significant effect size for binge/purge AN. Although the meta-analysis by Wu et al. (2014) did include studies with children and adolescents, the results did not differentiate between age groups.

Within the ASD literature, five reviews have attempted to synthesize the literature on cognitive flexibility (Hill, 2004; Kenworthy et al. 2008; Sanders et al. 2008; Geurts et al. 2009; Leung & Zakzanis, 2014). Studies using the WCST have consistently shown difficulties with set-shifting which are stable over time (Hill, 2004); however, this review did not differentiate between studies with adult or child participants and no quantitative synthesis was attempted. A further review (Kenworthy et al. 2008) reports that whilst several studies of cognitive flexibility in children and adults with ASD have demonstrated difficulties relative to HCs, the evidence is inconclusive with two of nine studies using the WCST, one with adults and one with children, producing unclear results. A recent meta-analysis (Leung & Zakzanis, 2014) comparing neuropsychological performance between patients with ASD and HCs on number of perseverative errors on the computerized WCST, which again pooled both child and adult studies, reported a medium mean effect size ($d = 0.68$). A review of the WCST in ASD (Landry & Al-Taie, 2016) provided effect sizes for four different WCST outputs including perseveration. In this review, age was found to be negatively correlated with perseveration, the opposite of what has been observed in the AN literature. To our knowledge, however, no reviews have directly compared the set-shifting profiles of children and adults with ASD using the WCST, although, generally, results do suggest that difficulties in set-shifting may be more stable across different age groups, than in AN studies.

Despite the set-shifting difficulties found in people with AN seeming to resemble those of ASD, the cognitive profiles of these two disorders have rarely been directly compared. Oldershaw et al. (2011) statistically compared the neuropsychological profile of currently ill AN patients with published data of individuals with ASD on the WCST and found the two groups to be statistically similar. To our knowledge, this is the only study directly comparing the profiles of the two groups and thus further work is needed to confirm whether the set-shifting profile seen in AN truly resembles that of ASD. Furthermore, the Oldershaw et al. (2011) study only included adult patients. Exploring differences in the set-shifting profiles of children and adults with AN would speak to debates of whether neurocognitive traits such as set-shifting underlie the disorder. Since the publication of the study of Oldershaw et al. (2011), additional studies have been published in both AN and ASD, as demonstrated by the review presented here. As these studies are cross-sectional and not longitudinal, inferences on the aetiology of ASD and AN cannot be made but statistically comparing the severity of set-shifting difficulties will aid in the comparison of the neuropsychological profiles of the two disorders.

Difficulties with set-shifting have been implicated in several other psychiatric disorders including schizophrenia (Pantelis et al. 1999); major depression (Austin et al. 1999) and obsessive–compulsive disorder (Leopald & Backenstrass, 2015), suggesting that this could be a general risk factor for the development of mental disorder, rather than a trait, specific to either AN or ASD. However, evidence of deviant set-shifting in these disorders is less consistent than in AN and ASD, for example, due to the heterogeneous nature of the disorders. As specific attention has been paid to the shared behavioural traits and intermediate phenotypes in ASD and AN (Treasure, 2013), determining
both the severity and stability of set-shifting difficulties in the two disorders through meta-analysis is a necessary step to exploring shared underlying mechanisms.

If it is the case that both children and adults with ASD display inefficiencies in set-shifting, it would add evidence to suggest that this trait is stable across the trajectory of the disorder. AN studies have tended to find less difficulty with set-shifting in children than in adults but this finding is difficult to interpret as it may reflect type II errors due to small sample sizes in child studies, with the exception of one study (Lang et al. 2015a). It is therefore unknown whether older and younger people with AN show similar cognitive profiles. The present review aims to use sufficiently powered analysis using pooled previous samples to summarize the literature on the use of the WCST as a measure of set-shifting in both ASD and AN populations and to determine: (a) whether individuals with either ASD or AN have difficulties with set-shifting; and (b) whether the severity of these difficulties is comparable in the two disorders in both child and adult populations. This will move us closer to understanding whether the difficulties in set-shifting are stable over the course of AN and will be a necessary but not sufficient step in understanding the relationship between AN and ASD.

Information sources and search

The electronic databases Scopus, PsycINFO, PubMed and Web of Science were searched up to and including January 2016. Search terms included anorexia nervosa OR autism AND set-shifting, Wisconsin, executive function OR cognitive flexibility. Limits included English-language, peer-reviewed articles. Reference lists of eligible papers were also screened, as were existing systematic reviews related of set-shifting in either ASD or AN (Roberts et al. 2007; Lang et al. 2014; Tchanturia, 2015; Landry & Al-Taie, 2016).

Selection

The first and principal authors (H.W., K.T.) identified potential titles and screened the abstracts for relevance. Full texts were then retrieved and read to determine eligibility by H.W. Texts deemed eligible were then further screened by K.T. and any papers that did not meet inclusion criteria were excluded. Reasons for exclusion included: the WCST was not used; perseverative errors were not reported; the WCST had been adapted for use with a particular population or there was no HC group; recovered AN cases were also excluded.

Data collection and items

The data items collected from each eligible study were: diagnosis; number of participants; mean age of participants; WCST version; percentage of female participants; how HCs were matched to clinical samples; how co-morbidities were controlled for; diagnostic tools; participant intelligence quotient (IQ); country where the study was conducted; and source of participant recruitment. For AN studies, body mass index (BMI) and illness duration were also extracted. In addition to demographic and experimental paradigm data, the means and standard deviations of perseverative errors on the WCST were used within the meta-analysis.

Risk of bias in individual studies

The risk of bias in individual studies was assessed by considering how methodology would make an impact on effect size in each study, for example, by attending to how HCs were matched to clinical samples, how the WCST was administered and how participants were recruited.

Summary measure

The principal summary measure used for analysis from all studies was the difference in means and standard deviations of perseverative errors or percentage of
Table 1. Demographic information for AN studies

<table>
<thead>
<tr>
<th>Author/date</th>
<th>Group</th>
<th>$n$</th>
<th>Mean age, years (S.D.)</th>
<th>Mean BMI, kg/m² (S.D.)</th>
<th>Mean illness duration, years (S.D.)</th>
<th>Mean IQ (S.D.)</th>
<th>HCs matched by</th>
<th>WCST version</th>
<th>% Female AN</th>
<th>% Female HC</th>
<th>Recruited from</th>
<th>Diagnostic tool</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbate-Daga et al. (2011)</td>
<td>AN-R</td>
<td>30</td>
<td>24.1 (6.7)</td>
<td>15.6 (1.6)</td>
<td>5.2 (4.2)</td>
<td></td>
<td>Age</td>
<td>Paper version</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 13</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>30</td>
<td>24.7 (2.6)</td>
<td>21.2 (2.2)</td>
<td></td>
<td></td>
<td>Race</td>
<td>128 cards, paper version</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 11</td>
</tr>
<tr>
<td>Aloi et al. (2015)</td>
<td>AN-R</td>
<td>45</td>
<td>22.8 (5.6)</td>
<td>15.5 (1.4)</td>
<td></td>
<td></td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>High school, university</td>
<td>Psychiatric hospital</td>
<td>SCID 13</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>45</td>
<td>25.6 (3.5)</td>
<td>20.2 (1.6)</td>
<td></td>
<td></td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 15</td>
</tr>
<tr>
<td>Fagundo et al. (2012)</td>
<td>AN</td>
<td>35</td>
<td>28.1 (8.2)</td>
<td>17.2 (1.4)</td>
<td></td>
<td></td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>Local community</td>
<td>ECIDE</td>
<td>SCID 13</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>137</td>
<td>24.8 (7.0)</td>
<td>21.5 (2.7)</td>
<td></td>
<td></td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 13</td>
</tr>
<tr>
<td>Fassino et al. (2002)</td>
<td>AN-R</td>
<td>20</td>
<td>23.8 (6.6)</td>
<td>15.2 (2.2)</td>
<td></td>
<td></td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>SCID 13</td>
<td></td>
</tr>
<tr>
<td>Galimberti et al. (2013)</td>
<td>AN</td>
<td>29</td>
<td>24.1 (6.8)</td>
<td>16.2 (4.0)</td>
<td>6 (5.1)</td>
<td></td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 15</td>
</tr>
<tr>
<td>Giel et al. (2012)</td>
<td>AN</td>
<td>15</td>
<td>23.9 (5.7)</td>
<td>15.8 (1.9)</td>
<td>103.0 (1.0)</td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>EDE</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>35</td>
<td>30.2 (10.9)</td>
<td>22.0 (1.9)</td>
<td>110.0 (17.0)</td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>0</td>
<td>0</td>
<td>ED service, mixed treatment</td>
<td>SCID 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goddard et al. (2014)</td>
<td>AN</td>
<td>29</td>
<td>26.2 (8.2)</td>
<td>17.7 (2.2)</td>
<td>106.9 (9.4)</td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>0</td>
<td>0</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 15</td>
<td></td>
</tr>
<tr>
<td>Harrison et al. (2012)</td>
<td>AN</td>
<td>50</td>
<td>27.1 (9.3)</td>
<td>15.4 (1.8)</td>
<td>108.4 (8.2)</td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>0</td>
<td>0</td>
<td>ED service, mixed treatment</td>
<td>University</td>
<td>SCID 13</td>
<td></td>
</tr>
<tr>
<td>Koba et al. (2002)</td>
<td>AN</td>
<td>11</td>
<td>28.5 (9.9)</td>
<td>21.6 (1.9)</td>
<td>112.3 (7.4)</td>
<td>Age, gender, education</td>
<td>Computerized</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>SCID 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakazato et al. (2010)</td>
<td>AN</td>
<td>27</td>
<td>27.7 (10.6)</td>
<td>15.4 (1.9)</td>
<td>10 (10.6)</td>
<td>Age, gender, education</td>
<td>Computerized, 4th edition</td>
<td>100</td>
<td>100</td>
<td>ED service, mixed treatment</td>
<td>SCID 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Group</td>
<td>Gender, Age, IQ, Class, Location</td>
<td>Test</td>
<td>Trials</td>
<td>Service, Treatment, Other Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------</td>
<td>-------------------------------------</td>
<td>------</td>
<td>--------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ohmann et al. (2004)</td>
<td>HC</td>
<td>28 26.9 (5.8) 22.3 (2.5) 15.2 (1.2) 5.5 (5.2)</td>
<td>Age, gender</td>
<td>Modified WCST</td>
<td>100 100 Local community Psychiatric in-patient service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pignatti &amp; Bernasconi (2013)</td>
<td>HC</td>
<td>11 27.5 (6.3) 21.3 (2.1) 16.0 (2.0)</td>
<td>Age, education</td>
<td>Italian version</td>
<td>100 100 ED service DSM-IV-TR 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sato et al. (2013)</td>
<td>HC</td>
<td>20 27.8 (7.0) 21.6 (1.4) 14.6 (1.5) 3.6 (3.7) 97.8 (13.7)</td>
<td>Age, IQ</td>
<td>Computerized 128 trials</td>
<td>100 100 ED service, mixed treatment DSM-IV-TR 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steinglass et al. (2006)</td>
<td>HC</td>
<td>15 22 (3.0) 20.6 (1.2) 19 (1.0) 10.8 (5.8) 109.8 (12.1)</td>
<td>Gender</td>
<td>Computerized, 4th edition</td>
<td>100 100 ED research clinic SCID 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tchanturia et al. (2012)</td>
<td>HC</td>
<td>11 24 (3.1) 22.1 (1.8) 15.2 (1.9)</td>
<td>Age</td>
<td>Computerized, 4th edition</td>
<td>100 100 ED service, mixed treatment DSM-IV 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thompson (1993)</td>
<td>HC</td>
<td>10 25.8</td>
<td>Age, education, IQ, class, location</td>
<td>100 100 ED out-patient DSM-III-R 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Autreve et al. (2013)</td>
<td>HC</td>
<td>10 23.2</td>
<td>119.7</td>
<td>Gender</td>
<td>Computerized, 4th edition</td>
<td>100 100 ED service, mixed treatment University DSM-IV 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child/adolescent studies</td>
<td>AN</td>
<td>37 15.4 (1.5) 15.3 (1.4) 13.2 (10.7)</td>
<td>Age</td>
<td>94.6 90.2</td>
<td>Psychiatric hospital DSM-IV-TR 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andres-Perpina et al. (2011)</td>
<td>HC</td>
<td>41 15.4 (1.5)</td>
<td>119.5 (14.5)</td>
<td>Computerized, 4th edition</td>
<td>100 100 Local school ED in-patient DSM-IV 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitzpatrick et al. (2012)</td>
<td>HC</td>
<td>32 14.9 (1.9) 78.1 (6.6)</td>
<td>105.7 (12.8)</td>
<td>Gender, age</td>
<td>Computerized, 4th edition</td>
<td>100 100 Local community ED service, mixed treatment SCID, DSM-V 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lang et al. (2015a)</td>
<td>AN</td>
<td>41 15.1 (1.8)</td>
<td>107.1 (13.8)</td>
<td>111.6 (10.7)</td>
<td>Age, SES, IQ, ethnicity Computerized</td>
<td>100 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McAnarney et al. (2011)</td>
<td>AN-R</td>
<td>24 16.3 (1.2) 16.7 (1.3)</td>
<td>111.6 (10.7)</td>
<td>109.5 (11.0)</td>
<td>107.1 (13.8)</td>
<td>105.7 (12.8)</td>
<td>Computerized, 4th edition</td>
<td>100 100 Mixed treatment University DSM-IV 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
perseverative errors between the clinical sample (ASD or AN) and HCs on the WCST. Standardized mean scores were used within the analysis to account for the variation in version of the WCST used in the studies. The WCST involves sorting cards into one of four categories according to one of three sorting rules: colour, shape or number. Once a card has been sorted, the participant is given feedback as to whether they sorted the card according to the correct rule. The paradigm yields a number of different outputs which can be analysed including perseverative errors, when the participant continues to sort cards according to a previous rule (Lezak, 1995). The WCST is a widely used measure of general executive function (Heaton et al. 1993; Greve et al. 2005). The WCST is assumed to consist of three factors: perseveration, failure to maintain set and idiosyncratic sorting (Greve et al. 1998). An increased number of perseverative errors has been associated with frontal lobe dysfunction (Milner, 1963; Nelson, 1976) and, as a robust measure of conceptual flexibility, was therefore the main focus for this review.

Only studies using the WCST, rather than other set-shifting paradigms, were included because this is a widely used measure across a wide range of disorders, believed to reliably assess cognitive flexibility (Holmén et al. 2012). There is less variability in terms of test properties (64- or 128-card version, computerized or pen-paper) compared with other paradigms such as the trail-making test (Atkinson & Ryan, 2008). Unlike other paradigms, the WCST also provides numerous different outputs, including perseveration, the most relevant measure in relation to set-shifting. It is widely used in both AN and ASD research allowing for direct comparison between these groups. The WCST is also considered to be a purer measure of set-shifting, compared with other paradigms which may tap into other cognitive domains (Steinglass et al. 2006).

Synthesis of data

Synthesis of results for meta-analysis meant that only studies reporting mean scores of perseverative errors were included in the analysis. Therefore, studies which reported standardized rather than raw scores or presented results visually were not included. The reasons studies were excluded are presented in Fig. 1. The meta-analyses were performed by pooling the standardized effect sizes of each study using a random-effects model. Random-effects models assume that as well as within-group variability caused by sampling variability of scores, mean effect size is also caused by differences between studies (Borenstein et al. 2009). If the between-study heterogeneity is present, the random-effects model results in estimates
### Table 2. Demographic information for ASD studies

| Author/date         | Group | n   | Mean age, years (s.d.) | Mean IQ (s.d.) | HCs matched by | WCST version | % Female ASD | % Female HC | Recruited from | Diagnostic tool | Quality score |
|---------------------|-------|-----|------------------------|----------------|----------------|---------------|--------------|-------------|--------------|----------------|---------------|---------------|
| **Adult studies**   |       |     |                        |                |                |               |              |             |              |                |               |
| Ambery et al. (2006) | AS    | 27  | 37.6 (14.6)            | VIQ and PIQ    | 128 cards      | 18.5          | 20           | Specialist assessment service | ICD-10, ADI, ADOS | 13            |
|                     | HC    | 20  | 33.5 (12.0)            |                |                |               |              |              | Out-patient service           | ADOS          | 14            |
| Eack et al. (2013)  | ASD   | 43  | 24.9 (5.8)             |                |                | 12            | 33           | Out-patient service           | ADI, ADOS     | 12            |
|                     | HC    | 24  | 26.3 (5.5)             |                |                |               |              |              | ADI, ADOS     |               |               |
| Goldstein et al. (2001) | HFA  | 103 | 18.2 (10.1)           | FIQ and VIQ    | 128 cards      | 13.6          | 10.7         | University recruitment programme | ADI, ADOS, clinical description | 10            |
|                     | HC    | 103 | 19.0 (10.1)           | FIQ, PIQ and VIQ |                | 5.3           | 4.2          | University recruitment programme | APA criteria | 11            |
|                     | HC    | 22  | 33.5 (14.5)           |                |                |               |              |              | University recruitment programme | ADI, ADOS, clinical description | 10            |
| Lopez et al. (2005) | AS    | 17  | 29                     |                |                | 17.6          | 35.3         | University recruitment programme | ADI, ADOS, GARS | 10            |
|                     | HC    | 17  | 29                     |                |                |               |              |              | University medical centre             | DSM-III       | 11            |
| Maes et al. (2011)  | ASD   | 17  | 45.2 (13.1)           | Computerized   | 23.5           | 21.1         | University centre                     | DSM-IV-TR, AQ, CARS | 13            |
|                     | HC    | 19  | 38.4 (14.8)           | Highest level of education | 128 cards | 0            | 0            | Nationwide register                 | DSM-IV-TR, AQ, PARS, AQ-J | 13            |
|                     | HC    | 22  | 24.3 (4.0)            | Age            | Computerized   | 13.6          | 26.7         | Out-patient service           | ADI-R, PARS, AQ-J | 13            |
|                     | HC    | 33  | 26.8 (9.9)            |                |                |               |              |              |                     |               |               |
| Author/date | Group | n     | Mean age, years (s.d.) | Mean IQ (s.d.) | HCs matched by | WCST version | % Female ASD | % Female HC | Recruited from | Diagnostic tool | Quality score |
|------------|-------|-------|------------------------|----------------|----------------|--------------|--------------|-------------|--------------|----------------|---------------|--------------|
| **Adolescent/adult studies** |       |       |                        |                |                |              |              |             |              |                |               |              |
| Minshew et al. (1992) | AS    | 15    | 21.1 (8.0)             | 95.7 (13.6)    | Age, FIQ, gender and race |              | 0            | 0           |              | ADI, ADOS      | 10            |
|                      | HC    | 15    | 21.3 (8.3)             | 96.5 (12.3)    |               |              |              |             |              | ADI, ADOS, clinical description | 12            |
| Minshew et al. (1997) | AS    | 33    | 20.9 (9.7)             | 100.9 (13.0)   |               |              | 12.2         | 12.2        |              | ADI, ADOS      | 10            |
|                      | HC    | 33    | 21.2 (10.0)            | 100.5 (11.7)   | Age, FIQ, gender and race | 128 cards   |              |             |              | ADI, ADOS, clinical description | 10            |
| Minshew et al. (2002) | HFA   | 90    | 21.4 (9.7)             | 98.0 (16.0)    |               |              |              |             |              | ADI, ADOS      | 10            |
|                      | HC    | 107   | 21.2 (9.8)             | 100.9 (13.4)   |               |              |              |             |              | ADI, ADOS, clinical description | 10            |
| **Child studies** |       |       |                        |                |                |              |              |             |              |                |               |              |
| Kado et al. (2012) | PDD   | 52    | 118.5 (29.0)           | 97.7 (10.1)    |                | Keio version | 25           | 21.2        |              | DSM-IV         | 10            |
|                      | HC    | 52    | 119.7 (33.5)           | 109 (11.5)     | IQ and age     | Computerized | 0            | 0           |              | ICD-10, ADI-R, ADOS | 12            |
| Kaland et al. (2008) | AS/HFA| 13    | 16.4 (2.8)             | 109.6 (10.4)   | Gender, age, VIP, PIQ | 128 cards   |              |             |              | School         | 10            |
| Lam & Yeung (2012)  | HC    | 13    | 15.6 (3.1)             | 109.6 (10.4)   |                |              |              |             |              | Child ASD service | 10            |
|                      | ASD   | 12    | 6.1 (8.2)              |                |                |              |              |             |              |                |               |              |
| Li et al. (2014)    | HFA   | 38    | 9.6 (2.3)              | 109.8 (18.9)   | Age and gender | Computerized | 21.1         | 21.1        |              | Specialist service | 8             |
| Pascualvaca et al. (1998) | ASD | 31    | 10.2 (1.5)             | 113.0 (15.1)   |                | Manual 128 cards | 34.8         | 34.8        |              | DSM-II-R, CARS | 15            |
|                      | ASD   | 23    | 8.7 (1.1)              | 77.6 (14.5)    | PIQ            | Manual 128 cards | 34.8         | 34.8        |              | DSM-II-R, CARS | 15            |
| Pooragha et al. (2013) | HFA  | 15    | 6.7 (1.8)              | 110.2 (10.2)   | Age and education |              | 6.7          | 15.4        |              | School         | 10            |
|                      | HC    | 15    | 10.1 (2.4)             | 101.1 (9.1)    |                |              |              |             |              | Hospital        | 10            |
| Shu et al. (2001)   | AS    | 26    | 8.0 (12.0)             | 90.0 (19.1)    | Age            | Computerized 128 cards | 5.9          | 48           |              | Autism society | 14            |
| Tsuchiya et al. (2005) | HFA  | 17    | 12.5 (4.3)             | 92.3 (7.5)     | Age            | Computerized 128 cards | 5.9          | 48           |              | School         | 10            |
with wider confidence intervals (CIs) than fixed-effects models.

**Statistical analysis**

Analysis was carried out in STATA 13 (StataCorp, USA) with the following user contributed commands `metan` (Bradburn et al. 1998), `metabias`, `metatrim` (Steichen, 1998) and `metareg` (Sharp, 1998). Cohen’s $d$ was used to estimate the effect sizes and is reported for all studies together with 95% CIs. The effect sizes were interpreted according to Cohen’s (1988) definitions of small ($>0.20 \leq 0.50$), medium ($>0.50 \leq 0.80$), large ($>0.80 \leq 1.30$) and very large ($>1.30$). Positive effect size indicates that the clinical group (ASD or AN) made more perseverative errors than HCs. A $p$ value of $<0.05$ indicates significant difference between groups.

**Risk of bias across studies**

Publication bias was assessed by visually evaluating funnel plots of each study’s standardized mean differences against its precision and formally by Egger’s test (Egger et al. 1997) to see if a study’s precision was related to effect size (Borenstein et al. 2009).

**Additional analysis**

Between-study heterogeneity was measured by calculating $I^2$ (Higgins et al. 2003) based on Cochran’s $Q$ test: measure of heterogeneity, $I^2 = 100\% \times (Q – df)/Q$, where $df$ is degrees of freedom. $I^2$ ranges between 0%, indicating no heterogeneity and 100%, indicating high heterogeneity. Child/adolescent studies were analysed separately as well as overall effect size being calculated to explore whether effect was related to participant age. In addition, meta-regression was used to test for differences between ASD and AN participants using diagnosis (AN or ASD) as the moderator.

**Results**

**Study selection**

In all, 24 studies in ASD and 22 in AN were included in the meta-analyses consisting of a total of 815 AN participants matched to 956 HCs and 739 ASD participants matched to 760 HCs. The selection process for both AN and ASD studies is shown in Fig. 1.

**Study characteristics**

All extracted information for AN and ASD studies included in the systematic reviews and meta-analyses is presented in Tables 1 and 2, respectively. Quality
ratings for each study ranged from 16 (highest possible score) to 8 (Li et al. 2014). This study included very little information on the recruitment of clinical participants or HCs but the results were sufficiently detailed for it to be included in the current review. Of the AN studies, the lowest score was 9 (McAnarney et al. 2011). This was presented as a brief report and thus methodological detail was missing. The results of this study were also unclear so the authors were contacted by K.T. to clarify before data were included in the meta-analysis. The full version of the paper by Koba et al. (2002) was unavailable to the authors at the time of this review; however, as it has been included in previous reviews (Roberts et al. 2007) it was included in the current analysis.

One ASD study (Goldstein et al. 2001) included both child and adult participants. Although the authors did split the ASD participants into ‘old’ and ‘young’ for
analysis, the control group was not split so the pooled results of children and adults, with a mean age of 18 years, were included in the adult meta-analysis. The majority of ASD studies reported IQ which ranged from 77 to 113.1 in the ASD group and 89 to 118 in HCs, indicating no intellectual disability. Nine AN studies also reported IQ, all of which fell within the normal range (97.8–119.5 for AN and 104.9–119.7 for HCs). Of the AN studies, all but two included only female participants, whereas one study included only males (Goddard et al. 2014) and one included over 90% of females (Andres-Perpina et al. 2011). In the ASD studies three studies included only males, with the other studies ranging from 10 to 48% female participants.

Within both clinical groups, diagnosis varied. For example, within the ASD studies eight papers reported using a diagnosis of ASD, seven used high functioning autism, seven used Asperger’s syndrome and one used pervasive developmental disorder. In the AN studies, 17 used a diagnosis of AN while five only included participants with restrictive subtype AN, one of which had a separate binge-purge subtype AN group. Data on mean BMI were also extracted from AN studies and ranged from 14.4 to 19.0 in the AN group. In the ASD studies, diagnosis was given using a variety of clinical measures; eight used both the Autism Diagnostic Interview and the Autism Diagnostic Observation Schedule; 12 used diagnostic criteria taken from the DSM and several self-report measures including the Childhood Autism Rating Scale, the Autism Spectrum Quotient and the Social Responsiveness Scale were also used. In the AN studies, all studies used diagnostic criteria from the DSM, either assessed directly or by using the Structured Clinical Interview for DSM Disorders (SCID; American Psychiatric Association, 2013). One study used the Eating Disorder Examination (Fairburn & Cooper, 1993). Several ASD and AN studies excluded HCs who had a history of psychiatric disorder but few studies reported co-morbidities, although several AN studies assessed for depression, anxiety and/or obsessive-compulsive disorder.

Sources of participant recruitment also differed across studies; in AN studies, 10 studies recruited their AN sample from specialist eating disorder services consisting of a mixture of out-, day- and in-patient treatment. Four studies did not state what type of treatment the participants were receiving; one recruited from an out-patient service; one from an in-patient service and three from psychiatric hospital settings. HCs were recruited from universities, schools or the local community. In nine of the ASD studies, ASD participants were recruited from a variety of specialist, out-patient or in-patient services for people with ASD; four from autism societies, national registers or research recruitment programmes and one from a specialist school. HCs were recruited from recruitment programmes, schools or the community. Of the ASD studies, 10 were conducted in the USA; four in Japan; three in the UK; two in China and one in the Netherlands, Denmark, Hong Kong, Iran and Taiwan, respectively. Six of the AN studies were conducted in Italy; six in the UK; three in the USA; two in Germany, Japan and Spain and one in Belgium.

Risk of bias

There was good symmetry within all funnel plots indicating no relationship between effect and study size and thus no publication bias. Egger’s test indicated no evidence of publication bias (p’s > 0.21). Analysis for correction of publication bias (trim-and-fill method) estimated two missing studies but this did not change the effect size considerably, so uncorrected data are presented here.

Synthesis of results

The forest plot of AN studies included in the meta-analysis is shown in Fig. 2. The random-effects analysis with a total sample size of 1731 (AN = 815, HC = 916) revealed a significant difference between AN and HC groups on the number of perseverative errors on the WCST (d = 0.43, 95% CI 0.26–0.61, z = 4.83, p ≤ 0.001). When adult AN studies were analysed separately, there was a significant difference between AN and HC groups with a medium effect size (d = 0.48, 95% CI 0.28–0.68, z = 4.70, p ≤ 0.001). When the four child and adolescent studies were analysed separately, there was no significant difference between AN and HC groups (d = 0.25, 95% CI −0.05 to 0.55, z = 1.66, p = 0.096).

The forest plot of ASD studies included in the meta-analysis is displayed in Fig. 3. The random-effects analysis with a total sample size of 1499 (ASD = 739, HC = 760) revealed a significant difference between ASD and HC groups on the number of perseverative errors obtained during the WCST (d = 0.65, 95% CI 0.51–0.80, z = 9.91, p ≤ 0.001). For adult studies, there was a medium effect size (d = 0.68, 95% CI 0.44–0.91, z = 5.68, p ≤ 0.001). When adolescent studies were analysed separately, the effect size was also medium (d = 0.51, 95% CI 0.21–0.82, z = 3.28, p = 0.001). Child studies produced a medium pooled effect size (d = 0.70, 95% CI 0.46–0.92, z = 5.86, p ≤ 0.001).

Additional analysis

There was evidence of moderate heterogeneity in both AN (I² = 63.1%) and ASD (I² = 37.0%) studies. To
examine the difference between ASD and AN groups, meta-regression was conducted with diagnosis as the moderator. Within adult studies, there was no effect of diagnosis and thus no significant difference between how ASD and AN groups performed on the WCST ($n = 28$, mean difference = -0.20, 95% CI = -0.17 to 0.56, $t = 1.12, p = 0.274$). There was an almost significant difference between the children and adolescent ASD and AN groups, with the ASD group scoring, on average, worse on the WCST than children and adolescents of the AN groups; however, this difference did not reach significance (mean difference = 0.44, 95% CI = -0.007 to 0.89, $n = 14, t = 2.13, p = 0.053$).

**Discussion**

The aim of this review was to synthesize the literature on the use of the WCST in both ASD and AN to examine whether there are differences between children and adults both within and between these clinical populations. There were a number of differences between studies including how a diagnosis of either AN or ASD was given, the service and country in which participants were recruited from, the experimental paradigm used and how HCs were matched to the clinical groups. These differences may have made an impact on the individual outcomes of each study; however, generally the quality of the studies was deemed sufficient for meaningful synthesis and analyses. Studies also differed on the version of the WCST used, for example 128 vs. 64 cards and computerized vs. pen-paper versions. This may have had an impact on the results of each study. Children with ASD have been found to perform better on the computerized version of the test, suggesting that the versions are not equivalent and may measure slightly different constructs (Ozonoff, 1995).

The meta-analysis indicated significant differences between participants with AN and HCs, with the AN group making significantly more perseverative errors during the WCST. When children and adults data were analysed separately, the difference between children with AN and HCs was not significant, while in adults the difference remained significant, with a

Fig. 2. Forest plot of mean perseverative error score on the Wisconsin Card Sorting Test: standardized mean effect size for differences (SMD) between anorexia nervosa and healthy controls. CI, Confidence interval.
medium effect size. This indicates that adults with AN may have greater difficulty with set-shifting than young people with the disorder. This finding supports and extends a previous systematic review and meta-analysis (Roberts et al. 2007) which showed that adults with AN had significant difficulties with set-shifting relative to HCs. The addition of thirteen studies since Roberts et al.’s (2007) meta-analysis has confirmed the original finding of a significant difference between adults with AN and HCs. The addition of Lang et al.’s (2015a) empirical study to the child literature indicates a small non-significant effect of AN on set-shifting, in line with a previous systematic review (Lang et al. 2014). This non-significant difference between AN and HC groups may be driven by a lack of statistical power with a limited number of studies conducted with small numbers of participants. The Lang et al. (2015a) study was designed to specifically replicate the methods used within adult studies and to address previous methodological shortcomings in the child literature and therefore adds strength to the current literature, despite the result remaining non-significant. The analysis presented here shows that young people with AN may have difficulties with set-shifting, but further studies are needed to determine whether these difficulties are significant. The effect size in adult studies is larger than that in child/adolescent studies and longitudinal studies are required to properly test whether there is a magnification of the effect over time, whereby poor set-shifting is a risk factor for AN, but is also intensified by the illness, thus acting as a precipitating and perpetuating factor within the disorder. Another possibility is that more severe set-shifting problems during childhood are predictors of the persistence of these problems into adulthood. Alternatively, the WCST may be more reliable and sensitive to problems when used with adults, thus yielding bigger, significant effect sizes in adults.

Meta-analysis of ASD studies indicated a different pattern to the AN studies. Adults with ASD made

Fig. 3. Forest plot of mean perseverative error score on the Wisconsin Card Sorting Test: standardized mean effect size for differences (SMD) between autism spectrum disorder and healthy controls. CI, Confidence interval.
significantly more perseverative errors on the WCST in comparison with HCs, indicating significant difficulties with set-shifting with a medium effect size. When child studies were analysed separately, the effect size was also medium, demonstrating that difficulties with set-shifting in ASD appear to be stable across the lifespan. This is a novel finding, as little attention has been paid in the ASD literature to WCST performance across the lifespan; however, these findings support previous work that suggests that deficits on the WCST are stable over time (Huizinga & Van Der Molen, 2007) and studies of executive functioning deficits in children with ASD (Pellicano, 2007; Kimhi et al. 2014). The recent meta-analysis by Landry & Al-Taie (2016) found age to be negatively correlated with set-shifting ability. While planned statistical analysis of our review did not specifically test this, CI for the child and adult studies overlap so we are unable to conclude that there is a significant difference between children and adults with ASD.

Despite difficulties in set-shifting being widely reported within the ASD literature, preserved performance relative to HCs has also been documented in some studies (Ozonoff & Jensen, 1999; Russell & Hill, 2001). Hill (2004) attributes this discrepancy to methodological inconsistencies amongst studies; however, as the meta-analysis performed here only included studies using the WCST, it appears to provide strong evidence than individuals with ASD have problems in set-shifting, at least as assessed using the WCST. This is the first meta-analysis to statistically compare both adult and child ASD studies using the WCST and suggests very little difference between the two groups.

The meta-regression, analysing the impact of diagnosis on set-shifting ability, found no significant difference between ASD and AN, in either the adult or child/adolescent groups, although there was a non-significant trend for children with ASD to perform worse than children with AN. This suggests that there may be differences in the development of set-shifting difficulties across ages between AN and ASD, although larger, stronger-powered longitudinal studies would be needed to confirm this.

It is important to consider the implications of these meta-analyses for both AN and ASD populations. Within AN, interventions like cognitive remediation therapy (CRT) which targets cognitive processes have been developed and empirically tested (Tchanturia et al. 2013a, 2014, 2015), with improvements in set-shifting as well as high acceptability and engagement in the treatment being reported. In spite of this success, the use of executive function training in ASD is limited, with a paucity of research in this area (De Vries et al. 2015). The few studies that have focused on flexibility training in ASD have yielded mixed results, with one study reporting improvements in flexibility (Kenworthy et al. 2014), one showing no improvement (Fisher & Happe, 2005) and one suggesting a trend towards improvement with intention-to-treat analysis (De Vries et al. 2015). These mixed findings may suggest that the type of flexibility training currently offered to individuals with ASD is not suitable for this population. Trials in CRT for ASD are in their infancy but the success of CRT in other disorders along with the observation that children with ASD show development in executive functioning (Pellicano, 2010) makes it a worthwhile focus for future research.

While this review suggests that difficulties with set-shifting are not directly comparable in ASD and AN, its impact on symptoms and wider functioning is still unknown. One of the main cognitive theories of ASD proposes that deficits in executive function underlie the disorder (Russell, 1997) and that these difficulties may also cause problems with social interaction and inflexible behaviour (Pellicano, 2007; Geurts et al. 2014). In AN, the cognitive–interpersonal maintenance model of AN (Treasure & Schmidt, 2013) postulates that difficulty with flexibility is a pre-disposing factor for AN, making individuals more susceptible to societal rules and physical appearance. Thus, set-shifting may underlie other symptoms associated with the disorder. The observation that difficulties with set-shifting may underlie other symptoms in both ASD and AN suggests that transdiagnostic interventions which target cognitive style rather than direct symptoms could improve functioning and engagement in other therapies, such as cognitive–behavioural approaches which require a more flexible thinking style (Treasure et al. 2015).

This comparison between AN and ASD studies should, however, be interpreted in light of study limitations. The majority of ASD studies only included male participants whereas most AN studies included only females. Thus, there may be important gender differences within each disorder which make an impact on set-shifting ability. For example, in one study adolescent females with ASD outperformed their male counterparts on a test of set-shifting ability (Bolte et al. 2011); so if females with ASD were directly compared with females with AN, the results may be different. As starvation is thought to confound set-shifting ability in AN, it is also important to consider variables such as BMI and illness duration on WCST performance, which were not analysed in this review. A study in healthy female participants found set-shifting ability to be impaired after fasting for a 16 h period (Bolton et al. 2014), supporting the notion that in females at least, set-shifting is worsened following starvation. Thus, the question of whether poor set-shifting...
Ability is a trait or state characteristic remains unanswered, despite the evidence from this review suggesting that this difficulty is present in young people with AN. In the future, population-based data, rather than data from clinical samples regarding set-shifting difficulties in those with either subclinical AN or ASD traits, may be useful to avoid bias. The impact of diagnosis was not investigated here and as the diagnostic criteria of ASD and AN have become broader under the revised DSM-V criteria (American Psychiatric Association, 2013), it is possible that greater heterogeneity between studies exists, making studies less comparable.

Acknowledgements

H.W. and K.T. would like to thank the Medical Research Council, Swiss Anorexia Foundation and the Psychiatry Research Trust for their support. D.S. received financial support from the National Institute of Health Research Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, Psychology & Neuroscience, King’s College London.

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Declaration of Interest

None.

References


Galimberti E, Fadda E, Cavallini MC, Martoni RM, Erzegovesi S, Bellodi L (2013). Executive functioning in anorexia nervosa patients and their unaffected relatives. Psychiatry Research 208, 238–244.


