ORIGINAL PAPER

Are We Under-Estimating the Association Between Autism Symptoms?: The Importance of Considering Simultaneous Selection When Using Samples of Individuals Who Meet Diagnostic Criteria for an Autism Spectrum Disorder

Aja Louise Murray · Karen McKenzie ·
Renate Kuenssberg · Michael O'Donnell

Abstract The magnitude of symptom inter-correlations in diagnosed individuals has contributed to the evidence that autism spectrum disorders (ASD) is a fractionable disorder. Such correlations may substantially under-estimate the population correlations among symptoms due to simultaneous selection on the areas of deficit required for diagnosis. Using statistical simulations of this selection mechanism, we provide estimates of the extent of this bias, given different levels of population correlation between symptoms. We then use real data to compare domain inter-correlations in the Autism Spectrum Quotient, in those with ASD versus a combined ASD and non-ASD sample. Results from both studies indicate that samples restricted to individuals with a diagnosis of ASD potentially substantially under-estimate the magnitude of association between features of ASD.

Introduction

There is considerable interest in establishing whether symptoms of autism spectrum disorders (ASD) can be considered a coherent category of impairments underpinned by a common cause rather than the result of the confluence of multiple symptoms with distinct causal roots (e.g. Williams and Bowler 2014). The question has formed the basis of numerous empirical and review studies, including a recent special issue of the journal *Autism* (2014, vol 18, issue 1). Typically, these discussions are framed in terms of three core behavioural symptoms of autism referred to as the 'triad of autism' and which have, until recently, formed the basis of its diagnosis. These symptoms are deficits in reciprocal social interaction (Soc), communication (Comm), and restrictive and repetitive stereotyped behaviour [RSB: American Psychiatric Association (APA) 1994]. In the present study, we argue that evidence on the extent to which symptoms of autism are inter-related must take into account the issue of simultaneous selection on ASD symptoms. To ignore simultaneous selection is to potentially substantially underestimate the degree to which ASD symptoms tend to co-occur.

It has long been acknowledged that ASD is an extremely heterogeneous disorder (Rutter 2014), however, increasingly these observations have crystallised into the hypothesis that ASD is a fractionable disorder comprising multiple, somewhat independent, symptom domains (see

A. L. Murray (✉)
Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
e-mail: s0785823@staffmail.ed.ac.uk

A. L. Murray
Department of Psychology, University of Edinburgh, Edinburgh, UK

K. McKenzie
Community Child Health, NHS Lothian, Edinburgh, UK

R. Kuenssberg
NHS Fife, Kirkcaldy, UK

M. O'Donnell
School of Psychology, University of Ulster at Magee, Londonderry, UK

Happé and Ronald 2008 for a review). Specifically, the 'fractionable triad' hypothesis suggests that the three classical symptoms of ASD: deficits in reciprocal social interaction, communication, and restrictive and repetitive stereotyped behaviour, are not all manifestations of the same underlying disorder; but rather separate domains of impairment whose co-occurrence is what defines an individual suffering from ASD. Discussions have also expanded the hypothesis beyond the classical triad to consider the fractionation of ASD symptomology more broadly. For example, the DSM 5 (APA 2013) collapses the reciprocal social interaction and communication domains into a single social communication domain. Recent studies have, therefore, also considered fractionation in terms of these two domains (Mandy et al. 2014). Other studies have considered fractionation in terms of cognitive symptoms (Brunsdon and Happé 2014) or in terms of genetic and environmental etiology (Dworzynski et al. 2009; Mazefsky et al. 2008; Robinson et al. 2012).

It has been argued that, if the symptoms of ASD are fractionable, then this has some important substantive and practical implications. This may explain the considerable attention that the hypothesis has received in the literature. First, it multiplies the importance of taking care to achieve adequate coverage of all symptom domains during assessment because if ASD symptoms are relatively independent, then global assessments of ASD may fail to capture key features of an individual's symptom profile (Happé and Ronald 2008). Second, it implies that there is no requirement for ASD symptoms to be specific to ASD because, under the fractionable triad hypothesis, ASD is merely the co-occurrence rather than the cause of specific ASD symptoms. Third, distinct etiologies of ASD symptoms suggest that searches for specific causes should focus efforts on specific symptoms. A fourth possibility is that treatments will have symptom specific rather than global effects and, by the same token, should be targeted at specific symptoms to maximise chances of success.

Historically a key piece of evidence contributing to development and then subsequent testing of the fractionable triad hypothesis is the extent to which symptoms of ASD tend to correlate with one another in individuals with a clinical diagnosis of ASD (Happé and Ronald 2008). For example, several studies which have found only modest correlations among different ASD symptoms in individuals who meet the diagnostic criteria for ASD have been cited as evidence for a fractionable disorder (e.g. Brunsdon and Happé 2014; Dworzynski et al. 2009; Kolevzon et al. 2004; Mandy et al. 2014). It is important to consider, however, that low symptom inter-correlations in samples of clinically diagnosed individuals do not necessarily imply that the symptoms are not strongly inter-related in actuality. In the section that follows, we describe how range restriction may lead to substantial under-estimates of symptom inter-correlations in individuals who meet diagnostic criteria for ASD.

Individuals who meet the diagnostic criteria for ASD are a select group comprising approximately only 1 % of the population (Baird et al. 2006; Baron-Cohen et al. 2009). Importantly, such individuals are not a random sub-sample, but rather a select sub-section of the population representing the extremes of ASD traits. It has long been known that when samples are selected with respect to some trait, the variance of that trait and its correlation with other variables is attenuated (e.g. Pearson 1903). This is an issue known as 'range restriction' and it comes in many forms (Sackett and Yang 2000). The simplest form of range restriction is explicit or direct selection on some variable X, when the correlation between X and some other variable Y is of interest. That is, the variable X on which the sample is selected (the 'selection variable') is identical with the variable X which is utilised in analyses in the selected sample (the 'substantive variable'). Direct range restriction would occur if a researcher administered a test of ASD symptomology to a group of individuals and then proceeded to analyse the correlation between the scores from that test with some other criterion variable e.g. executive functioning, in only the subset of individuals who crossed the clinical threshold on the ASD test. Direct range restriction sometimes arises when X is some aptitude test used to select candidates for a job and Y is a measure of job performance administered to the successful candidates in order to validate the aptitude test X (e.g. Berry et al., 2011) but is otherwise unusual. Unfortunately, most cases of range restriction are more complex and cannot be adequately dealt with using the simple correction formula developed to correct for the effects of direct univariate selection (e.g. see Sackett and Yang 2000).

The process of diagnosing ASD is an example of a situation involving more complex range restriction. It, too, represents a selection process that reduces the variance in the traits of interest in the diagnosed population and which will, in turn, attenuate symptoms relative to the overall population but there are some additional factors to consider. First, ASD diagnosis does not involve direct selection on measured scores on the X. That is, a clinician cannot simply 'read off' an individual's level of Soc, Comm and RSB and assign those with a score above certain cut-off points on all three a diagnosis of ASD. Instead, the process involves a combination of formal assessment and clinical judgement which can vary widely across individuals (Allison et al. 2012). The upshot of this is that scores on measures of ASD symptoms obtained in a subsequent research study will not be completely identical to the criteria by which a clinician has assigned a diagnosis, even if the same test scores contribute in both instances.

The process of diagnosing ASD and then selecting participants for a research study, therefore, represents an example of indirect selection. Indirect selection is defined as the case when the selection variables are not identical with the substantive variables that form the basis of subsequent empirical analyses. The selection variables do, however, induce selection on the substantive variables because they correlate with these substantive variables (Hunter et al. 2006). In the terminology of range restriction, therefore, the triad or other features of ASD of interest in an empirical study represent 'incidental variables' which are selected by virtue of being correlated with the unmeasured variables on which selection (diagnosis) is truly taking place (i.e. the selection variables).

Another way in which the case of ASD diagnosis is complicated as an example of a situation involving range restriction is that, unlike the univariate example above, ASD diagnosis requires the presence of symptoms in several domains to be present in order for a diagnosis to be merited. The requirement for deficits in multiple areas means that ASD diagnosis represents a case of simultaneous selection (Sackett and Yang 2000). Under DSM IV criteria diagnosis required deficits in all of the Soc, Comm and RSB domains and was, therefore, a case of selection on three variables. As more cases become diagnosed under the new DSM 5 criteria (requiring deficits in Social Communication and RSB), ASD diagnosis will shift to a process of selection on two variables. A multivariate selection formula was developed by Aitken (1935) and subsequently extended by Lawley (1944) to deal with situations such as this in which samples are selected on multiple variables (see Supplementary materials). Based on this formula it would, in principle, be possible to obtain estimates of the correlation between ASD symptoms correcting for range restriction, however, this is not true in practice. This is because in order to apply this correction, it is necessary to have information on the selection variables that is simply not available in the case of ASD diagnosis because, as mentioned above, the selection variables are a composite of formal assessment and clinical judgement and the latter is not directly quantifiable. In fact, this information is rarely available for any multivariate selection problem (Hunter et al. 2006). This creates a challenge with respect to estimating the degree to which symptoms of ASD cluster together because any sample restricted to individuals with ASD will be liable to under-estimate their association. Further, owing to a lack of information on the selection variables, it will be difficult to assess the extent of the bias.

While the possibility that range restriction may undermine the validity of results from clinical ASD samples has been noted (Happé and Ronald 2008), there has not as yet been any systematic study of the consequences of this kind of selection. This is important because ASD is fundamentally a clinical disorder and inferences regarding ASD should, therefore, come at least in part from samples of individuals who are actually diagnosed with the disorder. Clearly, it would be undesirable to disregard all studies restricted to individuals with diagnosed ASD from consideration because they are affected by range restriction. It was, therefore, the aim of the present study to attempt to characterise and quantify the problem of simultaneous selection within research studies utilising individuals with a clinical diagnosis of ASD. We first present the results of a brief simulation exploring the potential effect of simultaneous selection on estimates of the inter-correlation among symptoms of ASD. We then provide a real data example comparing the correlation among ASD symptoms in individuals with a diagnosis of ASD to a combined sample which includes both individuals with and without a diagnosis of ASD.

Methods

Simulation Study

When an individual receives a diagnosis of ASD, there has traditionally been a requirement that they demonstrate deficits in all three domains of Soc, Comm and RSB. Therefore, the majority of samples of individuals with a clinical diagnosis of ASD are simultaneously selected on all of Soc, Comm and RSB. For the purposes of our simulation study, we, therefore, considered a triad of ASD symptoms because, while the new DSM 5 criteria require deficits in only two domains, the majority of studies to date have utilised participants diagnosed according to the three classical domains. We, therefore, explored the possible effect of simultaneous selection using a range of possible values for the population correlation between Soc, Comm and RSB. All analyses were conducted in R statistical software (R Core Team 2013).

We used a model in which Soc, Comm and RSB have a trivariate normal distribution with means of zero and variances of 1 in the population to generate data. This corresponds to the idea that the traits are normally distributed in the population, with ASD representing the extremes of these traits (e.g. Austin 2005; Lundström et al. 2012). We simulated 10,000,000 cases to represent the population and varied the population correlations between Soc, Comm and RSB across simulation conditions. The population correlations utilised are provided in Table 1. Reflecting the evidence that Comm and SS are more strongly inter-related than either is with RR, we simulated uneven population correlations among the triad (e.g. Dworzynski et al. 2009).

We then simulated simultaneous selection from these populations in a manner designed to represent the

Table 1 Extents of attenuation of symptom inter-correlations under simultaneous selection

RSB and Soc correlation			RSB and Comm correlation			Comm and Soc correlation		
Population correlation	Sample correlation	% bias	Population correlation	Sample correlation	% bias	Population correlation	Sample correlation	% bias
.70	.34	−51	.70	.35	−50	.95	.83	−13
.60	.26	−57	.60	.26	−57	.95	.83	−13
.50	.19	−62	.50	.20	−60	.95	.83	−13
.40	.13	−68	.40	.14	−65	.95	.83	−13
.30	.10	−67	.30	.10	−67	.95	.83	−13
.60	.26	−57	.60	.25	−58	.90	.70	−22
.50	.19	−62	.50	.19	−62	.90	.70	−22
.40	.13	−68	.40	.13	−68	.90	.71	−21
.30	.09	−70	.30	.09	−70	.90	.71	−21
.70	.35	−50	.70	.34	−51	.80	.49	−39
.60	.25	−58	.60	.24	−60	.80	.50	−38
.50	.18	−64	.50	.18	−64	.80	.51	−36
.40	.14	−65	.40	.14	−65	.80	.52	−35
.30	.10	−67	.30	.11	−63	.80	.51	−36
.60	.26	−57	.60	.26	−57	.70	.37	−47
.50	.19	−62	.50	.18	−64	.70	.37	−47
.40	.14	−65	.40	.14	−65	.70	.39	−44
.30	.10	−67	.30	.10	−67	.70	.39	−44
.50	.19	−62	.50	.19	−62	.60	.29	−52
.40	.14	−65	.40	.14	−65	.60	.29	−52
.30	.09	−70	.30	.10	−67	.60	.21	−65
.40	.14	−65	.40	.14	−65	.50	.21	−58
.30	.08	−73	.30	.09	−70	.50	.20	−60

diagnostic process by selecting cases from the uppermost part of the univariate distributions of the three variables. We did not select on the simulated Soc, Comm and RSB scores directly because in practice diagnoses are not made on the basis of the same scores that are examined in empirical studies of ASD. Rather, they are made on the basis of related criteria which are strongly predicted by, but not identical with the Soc, Comm and RSB themselves. To simulate this process we generated a 'selection variable' for each of the Soc, Comm and RSB variables. These selection variables were correlated at $r \approx .75$ with the corresponding symptom to represent this indirect selection. We selected cases based on being above the 95th percentile on these selection variables. The 95th percentile has previously been used to define abnormality in studies of ASD traits utilising general population participants (e.g. Robinson et al. 2012).

We examined the effect of simultaneous selection on the sample symptom inter-correlations. We quantified the degree to which these sample estimates under-estimate the corresponding population value using percentage bias, computed as:

$$(r' - r)/r \times 100\%$$

where r is the simulated population correlation and r' is the correlation in the selected sample.

Real Data Example

In our real data example we compared the correlations between ASD symptoms in clinically diagnosed individuals to the corresponding correlations in a combined sample of individuals with and without a diagnosis of ASD.

Measures

We utilised the Autism Spectrum Quotient (AQ: Baron-Cohen et al. 2001). The AQ is a 50 item questionnaire assessing 5 different domains: Social Skill, Attention Switching, Attention to Detail, Communication and Imagination. Half of the items are reverse keyed. For the current study, we scored the AQ on a dichotomous response scale resulting in a possible range of scores for each domain from 1 to 10. Item content is based on the classical triad of ASD as well as other cognitive traits associated with ASD. Therefore, with our real data example, we expand the consideration of symptom inter-

correlations beyond the classical triad to include other established features of ASD.

Numerous studies have suggested that the AQ has favourable psychometric properties including good test–retest reliability and acceptable internal consistency, higher scores in clinically diagnosed than control samples, normally distributed scores in the population and correlations with other features of ASD (e.g. Allison et al. 2012; Baron-Cohen et al. 2001; Takagishi et al. 2010). The advantage of the AQ in the context of the current study is that it is based on a dimensional approach to ASD which conceptualises ASD traits as continuous in the population and, therefore, measurable even in individuals who do not meet diagnostic criteria for ASD. Moreover, it was specifically designed to measure ASD traits across a broad range from normal to clinical levels. Indeed, evidence suggests that the AQ successfully captures variation in ASD traits in both clinically diagnosed and non-ASD individuals (Baron-Cohen et al. 2001; Hoekstra et al. 2011; Wheelwright et al. 2010).

Participants

Non-ASD Participants

Non-ASD participants came from 2 sources. First, 98 participants (27 males, 70 females and 1 'other' gender) came from an ongoing study of emotion recognition and ASD traits which included the AQ as a measure. The mean age of the sample was 31.0 (SD = 12.5). Participants were recruited online and from the university community, therefore, the vast majority of these participants reported their occupation as 'student'.

Second, 138 participants (27 males, 111 females) came from an ongoing study of sex differences in ASD traits. The mean age of the sample was 27.8 (SD = 12.5). Participants were recruited online and the sample, therefore, had a smaller proportion of students (n = 33) and a higher proportion of individuals from the wider population. Both of the studies above had received ethical approval from the first author's educational institution and participants gave informed consent to take part in the study.

ASD Participants

Participants with ASD came from a previous study of the AQ in clinically diagnosed individuals. The sample has been utilised and described in previous publications (Booth et al. 2013; Kuenssberg et al. 2014; Murray et al. 2014) and is described comprehensively in Kuenssberg et al. (2014). Ethical approval for the study was obtained from the local National Health Survey (NHS) ethics committee and Caldicott Guardians, and the relevant local institutional research department and data were collected retrospectively

from case files. The full sample includes 148 participants (107 males and 41 females) with a diagnosis of Asperger Syndrome or high functioning autism. High functioning autism was defined as meeting the criteria for autism but having normal intellectual functioning. Asperger syndrome was defined as meeting the criteria for high functioning autism but with no history of language delay. The mean age of the sample was 33.3 (SD = 10.7). In the current study, we used the subset of participants with complete data on the five domains measured by the AQ (N = 132–135). As the data were fully anonymised prior to receipt it is not possible to identify the specific demographic composition of this sub-sample.

Statistical Procedure

We computed an estimate of internal consistency for each of the AQ domains in the whole sample using Cronbach's alpha to obtain an estimate of scale reliability in the unselected population. We then computed Pearson correlations between AQ domain scores first in the ASD group and then in a sample that combined both the ASD and control participants. We quantified the difference in Pearson correlation between the whole sample and ASD sub-sample, in a similar way as in the simulation study by computing the percentage difference between whole and ASD sub-sample:

$$(r_{ASD} - r)/r \times 100\,\%$$

where r_{ASD} is the correlation in the ASD sub-sample and r is the correlation in the whole sample.

Results

Simulation Study

Results of simulating simultaneous selection on Soc, Comm and RSB are provided in Table 1. These show how a selection mechanism representing the ASD diagnosis can lead to substantial under-estimates of symptom inter-correlations in samples of clinically diagnosed individuals. For example, if our model of the selection mechanism is close to reality, then an observed correlation between RSB and Soc of $r = .26$ could belie a population correlation of $r = .60$. Other possible magnitudes of population and possible corresponding sample correlations can also be read off from Table 1.

The results also demonstrate how the biggest percentage under-estimates of the correlation among symptoms occur when the relevant population correlation is itself smaller (e.g. Taylor 2004). For example, the percentage bias for a population correlation of .95 was only −13 % whereas the

Table 2 Descriptive statistics for ASD, non-ASD and combined samples for the 5 AQ domains

Domain	Mean (SD) ASD sample	Mean (SD) non-ASD sample 1 (students)	Mean (SD) non-ASD sample 2 (web sample)	Mean (SD) combined non-ASD sample	Mean (SD) combined ASD and non-ASD sample	$\frac{SD_{ASD}}{SD_{Combined}}$
Social skills	7.93 (2.14)	2.94 (2.27)	3.91 (2.51)	3.51 (2.45)	5.12 (3.16)	.68
Attention switching	8.45 (1.51)	3.58 (2.19)	5.08 (2.37)	4.46 (2.41)	5.91 (2.87)	.53
Attention to detail	6.36 (2.30)	4.54 (2.36)	5.44 (2.21)	5.06 (2.31)	5.53 (2.39)	.96
Communication	7.16 (2.18)	2.15 (2.01)	3.51 (2.53)	2.95 (2.42)	4.48 (3.09)	.71
Imagination	6.10 (2.52)	2.16 (1.76)	2.70 (2.03)	2.47 (1.94)	3.80 (2.78)	.91

percentage bias for a population correlation of .40 was approximately −65 %. Therefore, to the extent that simultaneous selection attenuates symptom inter-correlations in ASD samples, it does so to a greater extent for those domains that have smaller population correlations.

Real Data Study

Descriptive statistics for the ASD, non-ASD and combined ASD and non-ASD samples are provided in Table 2. As expected, the mean scores for all 5 domains are higher in the ASD sample than in the non-ASD samples. The standard deviations do not differ markedly between the ASD and non-ASD groups but, as expected, were larger in the combined sample than in either of the ASD or non-ASD sub-samples. The SD ratios of the domain scores in the ASD to combined sample are provided in the last column of Table 2. The largest SD reduction was observed in the Attention Switching domain. The ratio of .53 for this domain is smaller than is often found in studies of reviews of range restriction ratios which have found approximate average ratios of .59 and ranges of .70–.91 depending on the substantive area studied (e.g. Alexander et al., 1989; Schmidt and Hunter 1977). The largest ratio was for the Attention to Detail domain (0.96) and suggested only minimal range restriction.

Cronbach's alpha values for the five domains were: Social Skills = .84, Attention Switching = .81, Attention to Detail = .67, Communication = .67 and Imagination = .78 estimated in the whole sample. Based on the ASD sample alone Cronbach's alpha levels were generally lower: Social Skills = .73, Attention Switching = .51, Attention to Detail = .68, Communication = .66 and Imagination = .72.

The correlations among symptom domains measured by the AQ in both the combined ASD and non-ASD sample and the ASD sub-sample are provided in Table 3. In the combined sample the correlations between symptom domains ranged from r = .32 to r = .82. With the exception of the Attention to Detail domain which did not

correlate well with other symptoms, all of the symptom correlations were large and >.65. In the ASD sub-sample, all of the symptom inter-correlations were substantially smaller than in the combined sample. In the ASD sub-sample, symptom inter-correlations ranged from r = .20 to .57. The percentage difference between the combined and ASD sub-sample ranged from −16 to −45 %. Therefore, the real data analysis supports the hypothesis that samples restricted to clinically diagnosed individuals could substantially under-estimate symptom inter-correlations.

Discussion

In the present study, we demonstrated that the selection process entailed in diagnosing individuals with ASD may lead to substantial attenuations of symptom inter-correlations. In addition, we presented evidence that, considering individuals with and without ASD together, the correlations among symptom domains can be quite large, even when only modest in individuals with a clinical diagnosis of ASD. This has implications for the hypothesis that ASD comprises relatively distinct symptoms because it suggests that previous studies utilising clinical samples could have underestimated the extent to which ASD symptoms correlate with one another.

In a brief simulation study we demonstrated the kinds of observed correlations between ASD symptoms that could be expected, given different levels of population inter-correlation and a selection mechanism corresponding to ASD diagnosis. These showed that symptom inter-correlations are potentially substantially reduced in samples restricted to individuals who meet diagnostic criteria for ASD. We complemented our simulation analysis with an examination of the difference between symptom inter-correlations in an ASD and a combined ASD and non-ASD sample in the five symptom domains measured by the AQ. In every case the correlation in the combined sample was substantially larger than in the ASD sub-sample, supporting the hypothesis that including only individuals who

Table 3 Correlation matrix of the 5 AQ domains in ASD versus combined sample

	Social skills	Attention switching	Attention to detail	Communication	Imagination
Social skills	–	.42 (−45 %)	.24 (−27 %)	.57 (−30 %)	.50 (−29 %)
Attention switching	.76	–	.29 (−26 %)	.51 (−31 %)	.44 (−33 %)
Attention to detail	.33	.39	–	.32 (−16 %)	.20 (−38 %)
Communication	.82	.74	.38	–	.51 (−26 %)
Imagination	.70	.66	.32	.69	–

Combined sample is below the diagonal, ASD sample is above the diagonal with percentage difference in parentheses

meet diagnostic criteria for ASD is liable to result in an attenuation of symptom inter-correlations not only in principle but in practice.

Counter to expectations, the biggest differences between the ASD and combined sample were not necessarily in those domains with the lowest correlations in the whole sample, as would be predicted based on the fact that larger population correlations usually yield smaller attenuations with range restriction. For example, the biggest percentage difference was observed in the correlation between Social Skills and Attention Switching and this correlation was large ($r = .76$) in the combined sample. Conversely, the smallest percentage difference was observed in the correlation between Communication and Attention to Detail which was modest in the combined sample ($r = .38$). One possibility is that some symptoms of ASD are more strongly selected than others during the diagnostic process. This would be consistent with the fact that standard deviation ratios in the five domains ranged all the way from .53 up to .96. Another possibility, however, is that the deviations of percentage biases from their predicted ordering across the domains is due to different levels of measurement error in the domains. Under indirect range restriction such as occurs in the case of ASD diagnosis, selection is related to the true scores of substantive variables and only indirectly to observed scores through true scores (Hunter et al. 2006). The degree of range restriction on observed scores u_X is then given by

$$u_X = \sqrt{r_{XX}u_T^2 - r_{XX} + 1}$$

where u_T is the degree of range restriction on X and r_{XX} is the reliability of X in the population. Therefore, for tests with lower reliability u_X is larger for the same u_T. The u_X for Attention to Detail was very large at .96 suggesting almost no range restriction, however, the reliability for this domain based on the combined sample was the smallest of all AQ domains at .67. On the other hand, u_X was very small for Attention Switching at .53 suggesting a high degree of range restriction, however, the internal consistency of this domain was also higher at .81. Therefore, the deviation of percentage bias magnitudes from expectations

based on combined sample correlations may partly reflect differential reliability.

Our simulation study focussed on the triad of ASD because it is within this framework that a large amount of the work on assessing the degree of fractionation of ASD symptoms has been conducted. However, similar considerations apply to other frameworks or features of ASD such as 'the dyad' of ASD or performance on theory of mind or executive function tasks. The extent to which the inter-correlations among these ASD features in clinically diagnosed samples are downwardly biased will depend on several factors. First, it will depend on the population correlation between the features of interest. The larger the population correlation, the smaller the attenuation of their association in a selected sample (Taylor 2004). In the case of the classical triad, this could accentuate the difference in inter-correlation between the Soc and Comm domains and the RSB domain that has led to the former two symptoms being combined into a single domain in DSM 5 criteria while the latter remains separated (e.g. Frazier et al. 2012).

Second, it will depend on how closely the symptoms of interest correspond to the variables on which a diagnosis has been made. In range restriction terminology, this is the degree of association between the selection variables and the incidental variables with stronger associations resulting in larger biases. For example, for individuals diagnosed on the basis of DSM IV, the correlations among the triad should be most strongly affected, with other features which are less directly selected on are less affected. However, as there is a heterogeneity in diagnostic methods across cases and clinicians, the association between the selection variables at diagnosis and the symptoms of interest in an empirical study utilising a diagnosed sample will not be uniform across all cases and this complicates the ability to predict the degree of association between the substantive variables of interest and the selection variables with any degree of precision.

Third, it will depend on how strong the selection mechanism is. While it is the goal of clinicians to successfully diagnose all individuals who genuinely meet the criteria for ASD and none who do not, uncertainty surrounding diagnosis is inevitable. Mis-classification of individuals with

respect to whether or not they receive a diagnosis of ASD can potentially serve to mitigate the effects of selection on symptom inter-correlations by weakening the strength of selection on ASD symptoms. Mis-classification rates will, in turn, depend on the sensitivity and specificity of the particular measures and method used to diagnose individuals but again, it is known that these can vary widely across cases and settings. Similarly, if diagnostic criteria result in more individuals receiving a diagnosis of ASD, with corresponding increases in prevalence, then the effect of diagnosis on symptom inter-correlations will be reduced. It is anticipated that the move from DSM IV to DSM 5 diagnostic criteria may result in a slight reduction in the prevalence of ASD (Maenner et al. 2014), therefore, the effect of diagnosis on symptom correlations may increase in the future as criteria become more strict. Strength of selection is also affected by the fact that a minority of individuals with a diagnosis of ASD receive that diagnosis in spite of not meeting all diagnostic criteria. To the extent that these individuals are included in empirical studies, this can also serve to mitigate the effects of clinical diagnosis on symptom inter-correlations, by weakening the strength of selection on ASD symptoms. Related to these considerations, empirical studies that utilise individuals who are selected in some other way than via clinical diagnosis will be subject to different degrees of range restriction. For example, using clinically referred individuals or relatives of individuals with ASD will introduce weaker selection on ASD traits and, therefore, be less strongly affected by issues of range restriction. On the other hand, samples which are selected on more stringent criteria than clinical diagnosis e.g. exhibiting very high levels of ASD will be subject to greater biases due to range restriction.

Finally, as mentioned above, the extent of bias will depend on the population reliability of the measure used. Specifically, smaller reductions in variance of observed scores will occur for the same degree of range restriction on true scores when reliability is low.

In the current study, we focussed on the consequences of samples restricted to individuals who meet diagnostic criteria for ASD, however, a number of studies which have been cited in discussions of the fractionable triad utilise only participants who do not have a diagnosis of ASD. For example, low factor inter-correlations between the domains of the AQ have been reported in samples of undergraduate students (Austin 2005; Hoekstra et al. 2008). By not including individuals with an ASD, a portion of the trait distributions corresponding to clinical levels will not be adequately represented in the sample and the variance in the traits of interest will be reduced as a consequence. These studies may be just as prone to under-estimating the association among symptom domains as samples restricted to individuals with a clinical diagnosis of ASD.

We also did not explicitly consider substantive sources of variation in the correlation between ASD symptoms such as age or sex differences. Range restriction may make it difficult to identify such genuine moderators although there may be theoretical reasons to expect them to exist. For example, with regards to age differences, Hobson (2014) has proposed that to the extent that ASD is a coherent syndrome, it may be so due to a common final pathway in which multiple possible deficits lead ultimately over the course of development to a coherent set of impairments. This suggests that the correlation between ASD symptoms could increase with age as individuals- and symptoms- develop. Similarly, sex differences in the prevalence and manifestation of ASD symptoms have been identified and some explanations for these observations imply that symptom inter-correlations could also differ by sex (Lai et al. 2011; Rivet and Matson 2011). Consistent with this, there is some evidence to suggest that the genetic and environmental correlations among symptom domains vary across males and females (Robinson et al. 2012).

Finally, though we framed our demonstration in terms of symptom inter-correlations because they have acquired a level of substantive importance in the literature, the consequences of selection on ASD traits are also not limited to correlations. Other statistics that depend on the variance or inter-correlation of variables in the sample will also be affected. This includes, for example, the reliability of psychometric assessments (Fife et al. 2012), genetic and environmental variances and correlations (Dominicus et al. 2006), and factor model parameters (Muthén 1990). For example, (Muthén 1990) demonstrated how factor loadings and factor inter-correlations are downwardly biased when factor analysis is conducted in a sample subject to range restriction. Through its impact on these statistics and the substantive conclusions that follow from their use, range restriction can impact both on the theoretical understanding of ASD as well as how it is diagnosed and treated in practice. The latter is the case because empirical evidence ultimately feeds back into and influences practice in the clinic via evidence-based diagnostic criteria and guidelines.

Collectively, these considerations suggest that investigations of ASD symptom inter-correlations and related statistics should recruit and jointly analyse results from participants both with and without ASD (e.g. Constantino et al. 2004). This approach is justified if it is assumed that clinical ASD is merely the extreme end of traits that are continuously distributed in the population (e.g. Frazier et al. 2012). That is, the assumption is that autistic traits are meaningful in the general population and not qualitatively different from the traits expressed by individuals with a clinical diagnosis of ASD. Such a viewpoint is becoming increasingly accepted. However, this approach also

requires the practical issue of measuring ASD traits across clinically diagnosed and general populations to be addressed (Murray et al. 2014). As Happé and Ronald (2008) noted, measures of ASD have face validity in clinically diagnosed samples but when the same measures are administered to individuals without a clinical diagnosis of ASD, it is not clear how the resulting data relates to clinical ASD. Measures of the broader autism phenotype (BAP) or autistic-like traits (ALTs) that aim to capture sub-clinical variation in ASD traits may be advantageous in this regard because they explicitly aim to capture levels of ASD traits that span normality and clinical ranges of the traits (see Wheelwright et al. 2010).

Limitations

It is important to consider the potential limitations of the current study. First, the simulation study was designed to reflect the process of diagnosis of ASD, however, because this selection cannot be characterised exactly, the possibility remains that the simulated process does not accurately reflect selection processes in the real world in some way. Second, our study focussed exclusively on symptom inter-correlations which have historically been important in the development of the fractionable triad hypothesis, however, these alone should not form the basis of substantive theory. For example, tests of the fractionable hypothesis have also considered other evidence such as conceptual analyses, genetic etiology and neural substrates of the cognitive features of ASD (e.g. Happé and Ronald 2008). Finally, our real data example was based on a convenience sample which is, therefore, not population representative. One possibility is that the correlations in the combined sample could over-estimate the population correlation if range enhancement has occurred as a result of this convenience sampling. The non-ASD sample, in particular included a disproportionate number of females and the ASD sample did not include any low functioning individuals. Therefore, these results should not be taken to be indicative of the 'true' correlation between ASD traits in the population, only as a demonstration that increasing the range of scores in the sample can increase the correlation between ASD traits. The sample size for the real data example was also relatively small and it will be an important future direction to replicate the study in samples which both approximate the underlying population better and are larger.

Conclusion

Samples restricted to individuals who meet the diagnostic criteria for ASD (or which exclude clinically diagnosed

individuals) are likely to substantially under-estimate the association between different symptoms of ASD as a result of range restriction. Given that substantive theories of ASD and the development of diagnostic and treatment processes may depend on the strength of inter-correlation between features of ASD, it will be important to take into account that observed associations in selected groups may not accurately reflect the association between features of ASD in the population.

References

Aitken, A. C. (1935). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society (Series 2), 4*(02), 106–110.

Alexander, R. A., Carson, K. P., Alliger, G. M., & Cronshaw, S. F. (1989). Empirical distributions of range restricted SD x in validity studies. *Journal of Applied Psychology, 74*(2), 253.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(2), 202–212.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences, 38*(2), 451–460.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., et al. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The Special Needs and Autism Project (SNAP). *The Lancet, 368*(9531), 210–215.

Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., et al. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *The British Journal of Psychiatry, 194*(6), 500–509.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.

Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology, 96*(5), 881–906.

Booth, T., Murray, A. L., McKenzie, K., Kuenssberg, R., O'Donnell, M., & Burnett, H. (2013). Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *Journal of Autism and Developmental Disorders, 43*(12), 2997–3000.

Brunsdon, V. E., & Happé, F. (2014). Exploring the 'fractionation' of autism at the cognitive level. *Autism, 18*(1), 17–30.

Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry, 45*(4), 719–726.

Dominicus, A., Palmgren, J., & Pedersen, N. L. (2006). Bias in variance components due to nonresponse in twin studies. *Twin Research and Human Genetics, 9*(2), 185–193.

Dworzynski, K., Happé, F., Bolton, P., & Ronald, A. (2009). Relationship between symptom domains in autism spectrum disorders: A population based twin study. *Journal of Autism and Developmental Disorders, 39*(8), 1197–1210.

Fife, D. A., Mendoza, J. L., & Terry, R. (2012). The assessment of reliability under range restriction: A comparison of α, ω, and test–retest reliability for dichotomous data. *Educational and Psychological Measurement, 72*(5), 862–888.

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., et al. (2012). Validation of proposed DSM 5 criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(1), 28–40.

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review, 18*(4), 287–304.

Hobson, R. P. (2014). The coherence of autism. *Autism, 18*(1), 6–16.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*(8), 1555–1566.

Hoekstra, R. A., Vinkhuyzen, A. A., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., et al. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-Short). *Journal of Autism and Developmental Disorders, 41*(5), 589–596.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*(3), 594.

Kolevzon, A., Smith, C. J., Schmeidler, J., Buxbaum, J. D., & Silverman, J. M. (2004). Familial symptom domains in monozygotic siblings with autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 129*(1), 76–81.

Kuenssberg, R., Murray, A. L., Booth, T., & McKenzie, K. (2014). Structural validation of the abridged Autism Spectrum Quotient-Short Form in a clinical sample of people with autism spectrum disorders. *Autism, 18*(2), 69–75.

Lai, M. C., Lombardo, M. V., Pasco, G., Ruigrok, A. N., Wheelwright, S. J., Sadek, S. A., et al. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PLoS ONE, 6*(6), e20835.

Lawley, D. N. (1944). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences, 62*(1), 28–30.

Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., et al. (2012). Autism spectrum disorders and autistic-like traits: Similar etiology in the extreme end and the normal variation. *Archives of General Psychiatry, 69*(1), 46–52.

Maenner, M. J., Rice, C. E., Arneson, C. L., Cunniff, C., Schieve, L. A., Carpenter, L. A., et al. (2014). Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. *JAMA Psychiatry, 71*(3), 292–300.

Mandy, W., Charman, T., Puura, K., & Skuse, D. (2014). Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: Evidence from Finnish and UK samples. *Autism, 18*(1), 45–54.

Mazefsky, C. A., Goin-Kochel, R. P., Riley, B. P., & Maes, H. H. (2008). Genetic and environmental influences on symptom domains in twins and siblings with autism. *Research in Autism Spectrum Disorders, 2*(2), 320–331.

Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014). Are Autistic traits measured equivalently in individuals with and without an Autism Spectrum Disorder? An invariance analysis of the Autism Spectrum Quotient Short Form. *Journal of Autism and Developmental Disorders, 44*(1), 55–64.

Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology, 43*(1), 131–143.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 200*, 1–66.

Rivet, T. T., & Matson, J. L. (2011). Review of gender differences in core symptomatology in autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(3), 957–976.

Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., et al. (2012). A multivariate twin study of autistic traits in 12-year-olds: Testing the fractionable autism triad hypothesis. *Behavior Genetics, 42*(2), 245–255.

Rutter, M. (2014). Addressing the issue of fractionation in autism spectrum disorder: A commentary on Brunsdon and Happé, Frazier et al., Hobson and Mandy et al. *Autism, 18*(1), 55–57.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112–118.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0, URL:http://www.R-project.org/.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*(5), 529–540.

Takagishi, H., Takahashi, T., Yamagishi, T., Shinada, M., Inukai, K., Tanida, S., et al. (2010). Salivary testosterone levels and autism-spectrum quotient in adults. *Neuroendocrinology Letters, 31*(6), 101–105.

Taylor, A. (2004). The consequences of selective participation on behavioral-genetic findings: Evidence from simulated and real data. *Twin Research, 7*(05), 485–504.

Wheelwright, S., Auyeung, B., Allison, C., & Baron-Cohen, S. (2010). Research defining the broader, medium and narrow autism phenotype among parents using the Autism Spectrum Quotient (AQ). *Molecular Autism, 1*(10), 1–9.

Williams, D. M., & Bowler, D. M. (2014). Autism spectrum disorder: Fractionable or coherent? *Autism, 18*(1), 2–5.