Designing and recruiting to UK autism spectrum disorder research databases: do they include representative children with valid ASD diagnoses?

F Warnell, B George, H McConachie, M Johnson, R Hardy, J R Parr

ABSTRACT

Objectives: (1) Describe how the Autism Spectrum Database-UK (ASD-UK) was established; (2) investigate the representativeness of the first 1000 children and families who participated, compared to those who chose not to; (3) investigate the reliability of the parent-reported Autism Spectrum Disorder (ASD) diagnoses, and present evidence about the validity of diagnoses, that is, whether children recruited actually have an ASD; (4) present evidence about the representativeness of the ASD-UK children and families, by comparing their characteristics with the first 1000 children and families from the regional Database of children with ASD living in the North East (Daslne), and children and families identified from epidemiological studies.

Setting: Recruitment through a network of 50 UK child health teams and self-referral.

Patients: Parents/carers with a child with ASD, aged 2–16 years, completed questionnaires about ASD and some gave professionals’ reports about their children.

Results: 1000 families registered with ASD-UK in 30 months. Children of families who participated, and of the 208 who chose not to, were found to be very similar on: gender ratio, year of birth, ASD diagnosis and social deprivation score. The reliability of parent-reported ASD diagnoses of children was very high when compared with clinical reports (over 96%); no database child without ASD was identified. A comparison of gender, ASD diagnosis, age at diagnosis, school placement, learning disability, and deprivation score of children and families from ASD-UK with 1084 children and families from Daslne, and families from population studies, showed that ASD-UK families are representative of families of children with ASD overall.

Conclusions: ASD-UK includes families providing parent-reported data about their child and family, who appear to be broadly representative of UK children with ASD. Families continue to join the databases and more than 3000 families can now be contacted by researchers about UK autism research.

Strengths and limitations of this study

- ASD is common, but researchers find recruitment to studies challenging. Families want to participate in ASD research; however, access to projects is often limited.
- The Database of children with ASD living in North East England (Daslne) has gathered information about a representative sample of children with ASD for over 10 years.
- Multidisciplinary team recruitment led to more than 1000 children and families being recruited to a new UK-wide Autism Spectrum Database (ASD-UK) over 30 months.
- There was no bias between families who consented and those who did not. Children and families were representative of UK children with ASD overall. Children’s ASD diagnoses were valid. One limitation is that, as parental consent is required, more disadvantaged groups are less likely to have participated.
- As of 2015, UK ASD databases include an unbiased cohort of over 3000 families who have provided data about their children, and have consented to be contacted for the ASD research.

INTRODUCTION

Autism Spectrum Disorders (ASD) affect around 1% of people, and have a significant impact on children and adults and their relatives. Recent data show that the life-time cost to the UK for an individual with ASD and intellectual disability (ID) is £1.5 million, and £0.92 million for ASD without ID. Research involving individuals with ASD and their families is, therefore, important to advance knowledge and improve treatment approaches and outcomes. However, despite ASD being common, researchers often find recruitment to studies challenging. A recent report shows that families are keen to...
participate in ASD research,\(^2\) but access to projects is often limited by location, and the lack of formal contact mechanisms. As a consequence, many studies recruit biased samples of participants, at a slower than expected rate, limiting the pace of research and potentially costing research funders more money.\(^3\) The challenge of recruiting sufficient participants may be solved through establishing large research databases or registries,\(^5\) so that families can be easily contacted about research that may interest them.

ASD research databases have already demonstrated their effectiveness. In the UK, McConachie and colleagues established the Database of children with ASD living in North East England (Dasl\(e\)e) in 2003. Dasl\(e\)e gathers information about children with ASD for service planning and facilitates contact with families about research. After 11 years, Dasl\(e\)e includes 1270 children and families, and has supported recruitment to 37 studies. Dasl\(e\)e’s partnership between health and education professionals led to the inclusion of around 55% of families within the geographical area covered.\(^5\)\(^6\)\(^7\) (McConachie H, personal communication, data from July 2014). McConachie and colleagues previously compared anonymised lists of responders with non-responders for gender and Townsend deprivation index, and showed that on these characteristics, Dasl\(e\)e families are broadly representative of families of children with ASD living in North East England.\(^9\)

Internationally, other ASD databases have been established. In the USA, the Interactive Autism Network (IAN) includes families who have a child with a clinical ASD diagnosis; once they have consented, families can be contacted about research studies. IAN includes over 45 000 people with ASD, and has supported over 300 research studies.\(^11\) However, the representativeness of IAN families compared with families of people with ASD generally is unknown. The Autism Spectrum Disorders, Canadian-American Research Consortium (ASD-CARCC), invites individuals with ASD and their parents from Canada and the US to complete a questionnaire, and consent to be contacted about research studies. Participants provide a diagnostic report as confirmation of their diagnosis.\(^12\)

Western Australia (WA) has a prospective population-based register, holding anonymous information on all people diagnosed with an ASD in WA since 1999 to provide information to assist in the planning for services and for research.\(^13\) However, of the new cases registered between 1999 and 2002, only 35% of parents gave consent to include identifying details; when clinicians encouraged participation, this rose to 56%.\(^14\)

In 2011, building on the success of Dasl\(e\)e, the national voluntary ASD research database for the UK—the Autism Spectrum Database-UK (ASD-UK)—was launched with the aim of: (1) including large numbers of children with ASD and families who were representative of UK families of children with ASD; (2) enabling researchers to have more access to families and (3) allowing families to participate more easily in ASD research.

This report describes the methods used to establish ASD-UK, the relationship with Dasl\(e\)e, and addresses the following research questions: (1) How similar are the children and families who participate, compared with those notified about ASD-UK but chose not to participate? (2) What is the reliability and validity of parent-reported ASD diagnoses? (3) To what extent are the characteristics of ASD-UK children and families similar to children and families from Dasl\(e\)e, and children and families recruited in population studies?

**METHODS**

ASD-UK and Dasl\(e\)e share similar recruitment methodologies and collect parallel data from different geographical areas of the UK; both databases are managed by the same team at Newcastle University.\(^9\) ASD-UK and Dasl\(e\)e operate as a single research database. No family is included on both databases.

**Recruitment to ASD-UK**

Following ethical and local approvals, recruitment of families started in April 2011 through a pilot network of 30 ‘research interested’ UK neurodisability, community child health and mental health teams who had expressed interest in the initiative; teams from the Dasl\(e\)e recruitment area were not able to participate. After the pilot phase, an additional 20 teams joined ASD-UK; these teams heard about the database at research meetings and showed interest in recruiting. These teams include around one-fifth of UK Child Development Teams.\(^15\)

Families with one or more children, aged 2–16 years, who were given a clinical ASD diagnosis were eligible for recruitment. ASD-UK participation was discussed at a diagnostic clinic appointment or subsequent appointment; clinicians wrote to families seen recently but who were no longer reviewed in the clinic. Families received an information sheet and Expression of Interest form (EoI); those who responded were telephoned by ASD-UK staff to explain the project; a pack was then sent that included a consent form, parent questionnaire and the Social Communication Questionnaire—lifetime version (SCQ);\(^16\) the SCQ was chosen to give some standardised information about ASD characteristics of the children. Some parents registered and completed the consent and parent questionnaire online (http://www.asd-uk.com). Families could also self-reference by contacting ASD-UK directly or via the website.

Families of UK children with an ASD diagnosis were invited to participate. Information collected from the parent questionnaire included the child’s date of birth, gender, name and address of clinician, ASD diagnosis, and a range of other information about their neurodevelopment (eg, the presence of coexisting conditions, and their learning and language ability); all questions from the existing Dasl\(e\)e questionnaire were included, with
exactly the same wording. Parents were invited to send copies of their child’s medical and educational report to enable validation of the ASD diagnosis and provide additional information about the child. Families who were approached but who chose not to participate were asked to send their child’s year of birth, gender, ethnicity, local authority and postcode; this information was sent to ASD-UK by some of these families. The last three digits of the postcode were deleted once a Townsend deprivation index score had been assigned. These data were collected for analysis of the representativeness of families who consented, compared to those who did not.

**Recruitment to Daslñe**

Daslñe’s aims, recruitment methods, and data about the validity of the ASD diagnoses of included children have been described previously. In brief, families were recruited from neurodisability, community child health and mental health teams, education and voluntary organisation sources in six local authority areas around Newcastle in North East England. Parents/carers living in these local authority areas were invited to join Daslñe shortly after their child received an ASD diagnosis; families from outside the area were unable to participate. Following informed consent, parents completed a parent questionnaire that included questions about their child’s neurodevelopment. Unlike ASD-UK, SCQ data were not collected from Daslñe parents. The child’s diagnostic status was validated by a questionnaire completed by their clinician. Capture-recapture methods were used to ensure that as many families as possible were approached about Daslñe, and ongoing close contact maintained with local clinicians so that coverage would be as complete as possible. Validation of children’s ASD diagnoses was previously examined by selecting 40 children at random and corroboration of diagnosis done by using standardised assessment measures or clinical notes.

**Analysis of ASD-UK data about representativeness and validity**

The representativeness of the families who consented to participate in ASD-UK was investigated in two ways. First, the characteristics of children from families who consented to join ASD-UK, and families who chose not to participate and returned data were compared (see above). Second, some parent-reported child and family characteristics were compared with Daslñe data, and other published data from population-based studies.

The investigation of the reliability and validity of the parent-reported ASD diagnosis was conducted in two ways: First, considering children for whom professional reports were available; every third child was selected in chronological order of recruitment. The parent-reported ASD diagnosis was compared with the diagnosis in professional reports. Second, when a professional report was available about a child, and the SCQ score was below 15 (giving a 0.85 sensitivity and 0.75 specificity for an ASD diagnosis), BG or JRP checked the reports for evidence that the child had ASD.

**RESULTS**

After 30 months recruitment, 1000 families had registered with ASD-UK. After 10 years recruitment, 1084 children were registered with Daslñe (approximately 55% of children with ASD from the area).

**Similarities between children from families who consented to join ASD-UK, and families who chose not to participate**

Data about families who chose not to participate in ASD-UK were available from 208 families; for these families, 6% of gender data and 7% of diagnosis data were missing. The number of families who received information about ASD-UK and who chose not to participate, but for whom basic data were not received, was unknown. For the consenting families, 2% of diagnosis data were missing. The characteristics of the children whose parents consented and those who did not consent but returned data were very similar for the available characteristics (table 1).

**Similarities between children from families who joined ASD-UK, and those who joined Daslñe, and to data from children and families from epidemiological studies**

Children from ASD-UK and Daslñe were similar for gender, age at diagnosis, type of ASD diagnosis, type of school attended, and proportion of children with a parent-reported learning disability. As expected from UK data, the mean Townsend deprivation score was higher for Daslñe in North East England than for ASD-UK families. (table 2)

Ethnicity data were available for 810 ASD-UK families (these data were not collected initially through ASD-UK), but not for families included in Daslñe. Eighty-nine percentage of parents reported their ethnicity was white; this number is very similar to the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Similarities between children from families who consented to join ASD-UK and families who chose not to participate, who provided the data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consenting families (n=1000)</td>
</tr>
<tr>
<td>Gender, %</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>80</td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
</tr>
<tr>
<td>Diagnosis, %</td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>20</td>
</tr>
<tr>
<td>ASD</td>
<td>60</td>
</tr>
<tr>
<td>Asperger syndrome</td>
<td>18</td>
</tr>
<tr>
<td>Median year of birth</td>
<td>2005</td>
</tr>
<tr>
<td>Mean Townsend score</td>
<td>0.86</td>
</tr>
</tbody>
</table>
the database data and SNAP diagnosis categories meant 54%, SNAP 55%). Methodological differences between (intellectual) disability (IQ <70: ASD-UK 48%, Daslne 49%); learning characteristics were similar: children’s gender (males: ASD-UK 80%, Daslne 85%, SNAP 72%); and learning (intellectual) disability (IQ <70: ASD-UK 48%, Daslne 54%, SNAP 55%). Methodological differences between the database data and SNAP diagnosis categories meant the proportions of children with ASD subtypes were not comparable (Baird et al reported the prevalence of children with ‘narrow autism’, ‘consensus autism’ and ‘other ASD’).

Reliability of parent and clinician-reported diagnoses, and validity of the ASD-UK autism spectrum diagnoses
For the 1000 ASD-UK children, 478 had professional reports available for checking; agreement between parent-reported and clinician-given diagnoses were investigated for 156 reports. In 142/156 professional reports, the ASD diagnosis matched the parent-reported diagnosis (91%). For 8/14 remaining children, the professional reports showed a difference between the professional and parent-reported diagnosis (eg, parents completed ‘Autism’ and the professional report stated ‘ASD’); in the context of autism as a spectrum diagnosis,20 parent-clinician reliability about diagnosis was 96%. Of the remaining six children, the report did not mention an ASD diagnosis specifically, but gave considerable information about the child’s social communication difficulties and repetitive behaviours. JRP reviewed the clinical reports for these six children and concluded there was sufficient evidence about ASD to corroborate the parent-reported diagnosis.

SCQ data were available on 94% of ASD-UK children (n=940); 90% scored ≥15. Of the 93 children with an SCQ score <15, 49 children had professional reports available, all of which either confirmed the child’s ASD diagnosis (44 children) or JP considered the information available in the reports was consistent with an ASD diagnosis (5 children).

**DISCUSSION**
ASD-UK and Daslne are linked databases of families of children with ASD who have indicated they are interested in research participation; each database recruits from different geographical areas but they have very similar methodologies. Some evidence that children from Daslne families are similar to children with ASD in the North East has been presented previously (McConachie et al, 2009). For the characteristics analysed, the children and families included on ASD-UK and Daslne are similar to one another, and are as expected for children with ASD and their families nationally1, 21 22 and not different from those notified who chose not to participate. Together, these data suggest that children and families are representative of UK families of children with ASD on some core characteristics. The reliability and validity of the parent-reported ASD diagnoses of children included on ASD-UK has been shown through comparison with clinical reports and scores on the SCQ. Overall, these data suggest that an unbiased cohort of families of children with reliable, valid diagnoses have provided data about their children, and can be contacted about UK ASD research projects. By July 2015, more than 3200 families had joined the combined databases. Thirty-nine studies have recruited through the databases (examples include core data analysis projects, surveys, cognitive neuroscience projects, intervention/treatment projects, research priority setting initiatives); many more studies will be supported in the coming years.

The reasons for the successful growth of ASD-UK include the partnership between researchers and UK teams who identify families; this has been fostered by visiting teams as they start recruiting, making presentations on progress at local/regional meetings, and sending professionals’ newsletters. As these recruiting teams are multidisciplinary and following, or working towards,
recently issued guidance about how best to diagnose children with ASD, it is unsurprising that the ASD diagnoses given by parents are valid and reassuring that we did not identify children without ASD on the database. These data are similar to the previous findings about the accuracy of parent-reported data, and a high ASD diagnosis validity rate from the IAN database. A second reason is that many parents are enthusiastic about participation in research. One example of a large study that recently recruited through ASD-UK/Dasl’s e is the ASD+ study of coexisting conditions in children with ASD (Koshy et al, in preparation); Koshy recruited 658 families in just 4 months—the largest UK study of children with ASD to date.

The challenges associated with gaining consent for the inclusion of identifiable data on research databases are well recognised. Workload burdens for clinicians and concerns about asking parents for consent at a sensitive time can affect recruitment. However, Mehta found that a clear, detailed explanation about the utility of a research database, and clear written information about the future uses of data led to effective recruitment. For ASD-UK, having a clinician discuss the database with each family, and giving clear information about the purpose of a research database has been a time-consuming but successful strategy. The collaboration with parents on designing appropriate database methods and clear materials is also likely to have been important. Regularly contacting families about research projects they can participate in and keeping them up-to-date with an annual newsletter that includes feedback from data given previously or studies that have recruited through the databases means that parents remain engaged and ready to respond to research requests.

Other methods of collecting data from large numbers of participants have been used successfully in autism research; so how does ASD-UK compare to these? A clinical database (ie, service users) may not include all cases in a given population and may be unrepresentative. In contrast, a register attempts to identify all cases of the condition/disease in the identified population and can, therefore, be used for epidemiological research, for example, a study linked population-based Swedish longitudinal register to explore the association between parental age and autism. The advantages of these population registers are that they are representative and provide information on large numbers of people for research. Studies undertaken through these registers can be done relatively quickly and cheaply. However, there are downsides, such as limitations on the quality and availability of much information as it has been derived primarily for administrative purposes. Furthermore, for information to enter the register, the patient/parent has to have contact with a doctor or hospital. On balance, while the method of including families adopted by ASD-UK does not lead to the very large sample sizes represented in these registers, by actively seeking family consent it addresses legal data protection and confidentiality requirements. This means the representative cohort of families can always be recontacted about studies or to give further data, unless they withdraw. For all databases where parents are contacted, non-response bias can affect utility. For example, Kalb et al achieved only a 28% response rate to an online survey sent to families registered with IAN. However, Koshy’s study recruited 40% of families, and showed no difference between the characteristics of responders and non-responders with the exception of a small difference in children’s age. (Koshy et al, in preparation)

There are some limitations to the ASD-UK recruitment method. Where consent for participation is required, more disadvantaged groups are likely to be underrepresented—the homeless, disabled, travellers, unemployed and people from more socially deprived backgrounds. There may also be an under-representation of children with a different first diagnosis. However, it is reassuring that our Townsend data showed little difference between the deprivation scores of those who consented and those who did not want to participate. We know that ASD-UK has an under-representation of families for whom English is not the first language; translating information into other languages is a future aim.

In the future, we anticipate ASD-UK will recruit from additional UK child health teams, and also recruit families through voluntary sector organisations. We are now recruiting a cohort of adults with ASD and relatives for a longitudinal study; older young people and relatives from the databases will be invited to join this cohort. Maintaining the representativeness of families while increasing the number of families who join the databases will be a main focus, as will supporting recruitment to a large number of ASD research studies.

Acknowledgements The authors would like to be grateful to the parents and children participating in ASD-UK and Dasl’s e, and the child-health teams who, through recruitment, have made a huge contribution to the success of the database. See www.asd-uk.com for a list of the participating NHS Trusts. They would also like to thank our colleagues who worked on the databases, and members of the Steering Committee, Research and Parent committees. They would particularly like to thank the parent representatives on these committees. They are grateful to Dr. Paul Norman, School of Geography, University of Leeds, who calculated the Townsend scores for the 2011 census wards.

Contributors JRP and HM lead ASD-UK and Dasl’s e respectively. FW was administrator for ASD-UK and MJ for Dasl’s e. RH manages the electronic databases and web-based data entry systems. FW, BG, JRP, MJ, and HM conducted data checking and analysis. FW wrote the manuscript with JRP and HM. All the authors had a role in contributing to the manuscript text. JRP is the guarantor.

Funding We are extremely grateful to the UK autism research charity Autistica (http://www.autistica.org.uk/) for funding and supporting ASD-UK and Dasl’s e. Dasl’s e has previously received funding from The Northern Rock Foundation for recruitment, and receives funding from The Children’s Foundation for database public engagement activities.

Competing interests None declared.

Ethics approval Newcastle and North Tyneside Research Ethics Committee 2.

Provenance and peer review Not commissioned; externally peer reviewed.
REFERENCES


