

Brief Report: Telephone Administration of the Autism Diagnostic Interview—Revised: Reliability and Suitability for Use in Research

Jessica Ward-King · Ira L. Cohen ·
Henderika Penning · Jeanette J. A. Holden

Published online: 2 March 2010
© Springer Science+Business Media, LLC 2010

Abstract The Autism Diagnostic Interview—Revised is one of the “gold standard” diagnostic tools for autism spectrum disorders. It is traditionally administered face-to-face. Cost and geographical concerns constrain the employment of the ADI-R for large-scale research projects. The telephone interview is a reasonable alternative, but has not yet been examined for reliability with face-to-face administration. In this study, participants were interviewed both face-to-face and on the telephone using the complete ADI-R interview. Results indicate that there was no significant difference between the algorithm scores or the diagnoses arrived at for face-to-face and telephone administrations. Reliability statistics across the two modalities were very good and indicate that telephone interviews using the ADI-R are a viable option for researchers.

Keywords Autism spectrum disorders · Autism diagnostic interview—revised · Telephone interview · Reliability

Introduction

The Autism Diagnostic Interview (ADI; Le Couteur et al. 1989) was originally intended for research, as an aid to diagnosing autism according to the ICD-10 (World Health Organization 1993) and DSM-IV (American Psychiatric Association 1994) definitions. This original version of the standardized, investigator-based interview was intended for caregivers of subjects with a chronological age of 5 years or older, and a mental age of at least 2 years. The interview is semi-structured, contributing to both its reliability and validity. It is based upon open-ended questions that inquire about various aspects of a subject’s development and current behaviour, allowing the informant to describe freely the relevant traits of the affected individual. The interviewer uses clearly defined codes to classify the traits and behaviours described in response to each probe.

The ADI-R fills a need in research for a sensitive and reliable tool based on accepted diagnostic criteria that can determine whether an individual fits into the diagnostic category of “autism” and has become a “gold standard” diagnostic tool for autism research. In the absence of a reliable tool, it is difficult for researchers to ensure consistent classification of individuals as having autism or not.

Since its creation in 1989, the ADI (and then the ADI-R) has been employed as a face-to-face semi-structured interview. For clinical use, this makes sense; the ADI-R is often used as part of a wider, multidisciplinary assessment in the

J. Ward-King · H. Penning · J. J. A. Holden
Department of Psychiatry, Queen’s University, Kingston,
ON K7M 8A6, Canada

J. J. A. Holden
Department of Physiology, Queen’s University, Kingston,
ON K7M 8A6, Canada

J. J. A. Holden (✉)
Genetics and Genomics Research Laboratory, Ongwanada,
Kingston, ON K7M 8A6, Canada
e-mail: holdenj@queensu.ca; holdenj@post.queensu.ca

I. L. Cohen
Department of Psychology, NYS Institute for Basic Research in
Developmental Disabilities, Staten Island, NY 10314, USA

J. Ward-King · I. L. Cohen · H. Penning · J. J. A. Holden
Autism Spectrum Disorders, Canadian-American Research
Consortium (ASD-CARC), www.AutismResearch.ca

diagnosis of autism, with families being seen in a clinic. For research purposes, the ADI-R is often used for diagnostic confirmation to empirically assess the previous clinical diagnosis that an individual has autism. It is not always feasible for a participating family to come to the research centre to see the interviewer or for an interviewer to travel to each family's home. Further, the use of the ADI-R for research requires that the interviewer obtain extensive training and must become reliable in scoring the interview with a designated research group (Le Couteur et al. 2009). It is difficult and expensive to train sufficient interviewers to send to each family's home, especially for large-scale studies that sample from a wide geographic area. In short, while the ADI-R has long been used successfully as a face-to-face interviewing tool, it has become necessary to adapt this reliable and valid diagnostic tool for the needs of large-scale research projects such as those designed to identify genetic and environmental factors leading to ASD susceptibility.

One group (Vrancic et al. 2002) adapted the ADI-R for administration over the telephone. In that study, the authors used the algorithm items of the ADI-R to develop an interview that consisted of 47 items and required approximately 20–40 min to complete. The wording of the interview was completely rephrased with necessary inclusion of examples and explanations to obtain reliable answers over the telephone. As it is not the complete ADI-R, the Autism Diagnostic Interview—Telephone Screening in Spanish (ADI-TSS) was developed as a screening tool that enables the selection of cases of probable autism. It was designed to compliment, and not replace, the ADI-R; the authors suggested that patients who are identified by the ADI-TSS should later be assessed using the ADI-R.

The ADI-R takes approximately 2 h to complete. Partly for this reason, the full interview has never been tested using the telephone modality. While the ADI-R can take several hours to complete, its authors report that informants find it an enlightening and comfortable experience “because they are allowed to describe important aspects of their child's behaviour in their own words” (Lord et al. 1994, p. 663). It is clear that it would be an advantage to autism spectrum disorder (ASD) research if the ADI-R could be administered in one session over the telephone, as it could then be used inexpensively and efficiently with minimum disruption for both the interviewers and participants. In this study we sought to determine whether it is possible for participants to complete the ADI-R with a trained ADI-R interviewer, and whether and how the telephone interview results would compare with face-to-face interview results in terms of reliability.

Methods

Participants

Twenty children with autism and their primary caregivers were recruited for this study. Children's ages at the time of the first interview ranged from 3.42 to 19.0 years (mean: 8.92 years). There were 14 boys and six girls included in the study, all of whom had a previous clinical diagnosis of an ASD; 15 with a diagnosis of Autistic Disorder; four with a diagnosis of Asperger's Disorder; and one with a diagnosis of PDD Not Otherwise Specified. No IQ or adaptive functioning data was available for the participants. One primary caregiver of each child (16 mothers, 4 fathers) volunteered to act as the informant on both face-to-face and telephone administrations of the ADI-R. All participants were recruited by the Autism Spectrum Disorders—Canadian and American Research Consortium (ASD-CARC) through an on-line Research Registry (www.AutismResearch.ca), which invites families to complete questionnaires online and to agree to be contacted when studies are being carried out in their area. For this study, participants were recruited from a circumscribed geographical area in Southeastern Ontario. All informants identified themselves as Caucasian.

Procedure

Informants were asked to participate in an interview using the Autism Diagnostic Interview—Revised (ADI-R; Lord et al. 1994) twice: once face-to-face with the interviewer, and once in an interviewer-initiated telephone call at a mutually convenient time. A single interviewer conducted all interviews. The order in which the interviews were conducted was counter-balanced. Interviews were completed at least 14 and not more than 122 days apart, with a mean interval of 29.6 days (standard deviation = 30.97 days), and all were completed within a 6-month period.

The interview was the complete, standard ADI-R interview (Lord et al. 1994) which includes items relating to both verbal and non-verbal individuals. The ADI-R was scored using the algorithm provided, but only at the conclusion of data collection, after all interviews had taken place. The interviewer (H.P.) was fully trained and certified in conducting and scoring the ADI-R for research purposes.

The ADI-R is scored using an algorithm that examines the main diagnostic criteria emphasized by the DSM-IV and ICD-10. There are four domains examined by the ADI-R. Domain A assesses qualitative abnormalities in reciprocal social interaction (QARSI), specifically examining use of eye-gaze and facial expression, development of peer relationships and emotional reciprocity, and seeking to

share one’s own enjoyment. Domain B assesses qualitative abnormalities in communication (QAC), and arrives at different scores for participants who are verbal (BV), and those who are non-verbal (BNV). For the present study, BV and BNV scores were analyzed together as one communication score since there were only four non-verbal participants and our concern was with repeat test reliability. Domain C assesses restricted, repetitive and stereotyped patterns of behaviour (RRSPB), including compulsions and unusual preoccupations as well as stereotyped mannerisms. Domain D deals with the requirement that abnormalities in the three diagnostic behavioural criteria be apparent before age 36 months.

Data Analyses

Domain means for the first three content domains were compared using repeated measures multivariate analysis of variance (MANOVA). Both domain and method of interview served as within-subjects factors. Order of the interview format (face to face or telephone) served as a between-subjects factor. Domain D scores at both interviews were negatively skewed (−1.5 and −1.2) and so these domain scores were compared across time with the Wilcoxon matched pairs test, and across group with the Mann–Whitney *U*, both non-parametric tests. Domain scores across interview conditions did not significantly

vary with prior diagnosis (all *p* values >0.17) and groups did not significantly differ in mean age ($F(1,18) = 0.7$, ns).

Results

Figure 1 shows the mean (±SEM) domain scores across interview conditions for the two groups. While those who received the telephone interview first had marginally lower group means on the first and second interviews, these differences were not statistically significant. Overall, there were no significant main effects for group ($F(1,18) = 1.07$, ns), method of interview ($F(1,18) = 0.07$, ns), or their interaction ($F(1,18) = 0.002$, ns). Multivariate tests also revealed no significant effects due to interaction of domain with group, domain by interview type, or their three-way interaction (all *p* values >0.6). Likewise, Domain D scores did not significantly vary across time or group (all *p* values >0.6). Table 1 shows the ADI-R means and standard deviations across interview conditions for the two groups.

Telephone interview domain scores were correlated (Pearson *R*) with face-to-face interview scores as the reference in order to examine repeat reliability. Correlations across domains A, B, C, and D were, respectively, 0.84, 0.73, 0.90, and 0.89; all *p* values <0.001. Mean (SD) ADI-R difference scores across subjects within conditions and interview conditions are shown in Table 2. As shown,

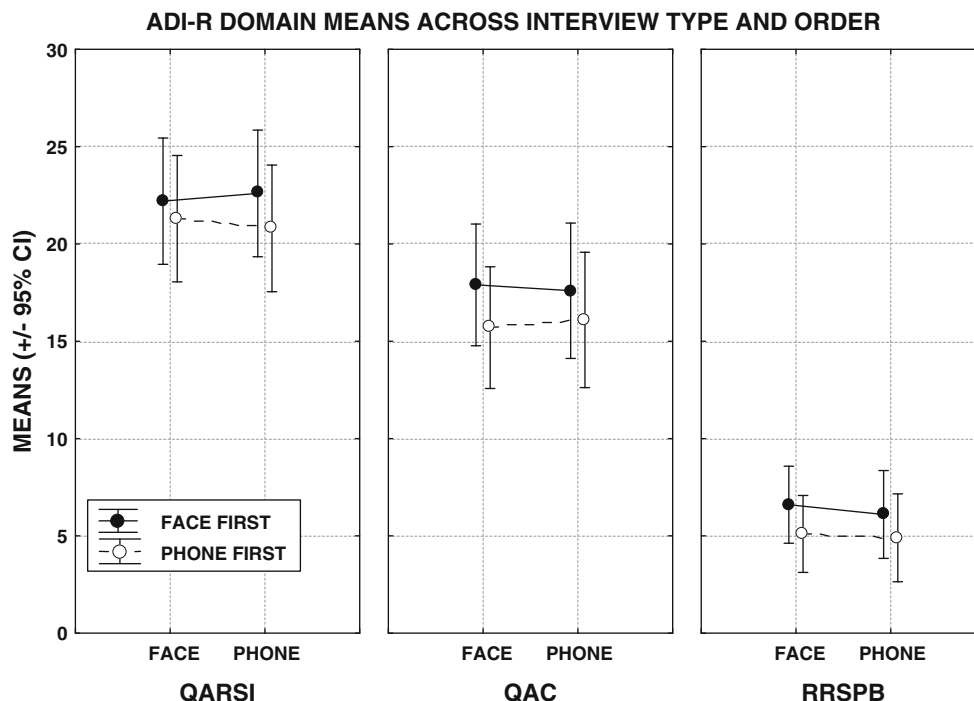


Fig. 1 ADI-R domain means (±95% confidence interval) across interview type and order. *QARSI* qualitative abnormalities in reciprocal social interaction, *QAC* qualitative abnormalities in communication, *RRSPB* restricted, repetitive and stereotyped patterns of behaviour

Table 1 Mean (SD) domain scores across group and method of interview for each ADI-R domain

| Group | N | Face | | | Telephone | | |
|-------------|----|--------------|--------------|-------------|--------------|--------------|-------------|
| | | QARS | QAC | RRSPB | QARS | QAC | RRSPB |
| Overall | 20 | 21.75 (4.78) | 16.80 (4.72) | 5.85 (3.00) | 21.70 (4.86) | 16.85 (5.15) | 5.50 (3.36) |
| Face first | 10 | 22.20 (4.37) | 17.90 (3.11) | 6.60 (3.37) | 22.60 (3.66) | 17.60 (5.04) | 6.10 (4.46) |
| Phone first | 10 | 21.30 (5.36) | 15.70 (5.89) | 5.10 (2.51) | 20.80 (5.88) | 16.10 (5.43) | 4.90 (1.80) |

Table 2 Mean (SD) difference scores across subjects within groups and method of interview for each ADI-R domain

| Group | N | QARS | QAC | RRSPB |
|--------------|----|--------------|--------------|-------------|
| Overall | 20 | 0.05 (2.72) | −0.05 (3.66) | 0.35 (1.50) |
| Face to face | 10 | −0.40 (3.03) | 0.30 (4.16) | 0.50 (1.58) |
| Telephone | 10 | 0.50 (2.46) | −0.40 (3.27) | 0.20 (1.48) |

difference scores did not significantly vary across groups and were close to zero ($F(2,36) = 0.5$, ns).

In terms of diagnostic agreement, 18 of the 20 cases met ADI-R criteria for autism based on the face-to-face interview. The other two cases met two of the three content domain criteria suggesting they could be classified as PDDNOS (Rutter et al. 2003). Of the 18 autism cases, 15 (83%) met criteria for autism using the telephone interview and the other 3 would have been classified as PDDNOS. Of the two cases classified as PDDNOS based on the face-to-face interview, one met criteria for autism and the other for PDDNOS using the telephone interview. Thus, irrespective of method, all cases remained in the autism spectrum. We did not compute Kappa for a measure of agreement because we did not have a non-spectrum comparison group; autism and PDDNOS groups are typically difficult to distinguish (Mahoney et al. 1998), and the number of PDDNOS cases was too small to provide a meaningful score.

Discussion

Our results indicate that the ADI-R remains a reliable diagnostic interview when it is administered over the telephone. There were no differences in the results, either on the diagnostic algorithm, or in terms of diagnosis reached depending on interview administration method.

The telephone has many advantages as an interview modality in research settings. For one, it is cost-effective (Bauman 1993; Burnard 1994; Corey and Freeman 1990; Marcus and Crane 1986; Musselwhite et al. 2007; Siemiatycki 1979; Tausig and Freeman 1988; Wilson et al. 1998); Marcus and Crane (1986) argue that telephone interviewing techniques could reduce costs 50–75% when compared to face-to-face interviews. Use of the telephone

to conduct interviews allows an interviewer to cover a larger geographical area (Burnard 1994; Musselwhite et al. 2007). Telephone interviews can be scheduled and completed more quickly than can face-to-face interviews (Worth and Tierney 1993).

There is scientific support for the telephone interview as a legitimate method of data collection (Oppenheim 1992; Barriball et al. 1996; Law 1997). Like face-to-face interviews, telephone interviews have a high response rate (Polit and Hungler 1991) and incorporate the possibility for the interviewer to clear up misunderstandings (Robson 1993). Robson (1993) also argues that they have smaller interviewer effects and a lower tendency for the respondent to give socially desirable responses. Quality control, which can be ensured with fewer, centrally-located interviewers who have the opportunity to self-correct, is a feature of telephone interviewing (Lavrakas 1987). The interviewers are also able to take interview notes more discreetly over the telephone, minimizing the discomfort that participants may have during the interview (Musselwhite et al. 2007).

Telephone interviews are not without disadvantages. Establishing rapport can be a problem in telephone interviews (Robson 1993). It is important to establish an appropriate relationship in order for the telephone interview to be successful, and for authentic responses to be provided. Furthermore, it has been argued that telephone interviews produce shorter responses than face-to-face interviews (Marcus and Crane 1986), possibly because they are more focused than face-to-face interviews (Carr and Worth 2001). The possible length of an interview changes with the modality, however. Many researchers advise a shorter telephone interview as compared to face-to-face interviews, with Lavrakas (1987) suggesting 20–30 min as a maximum. However, Waterman et al. (1999) found that telephone interviews of up to 60 min were “efficient in time and conducive to free-flowing conversation”. There is nothing preventing a telephone call from lasting for even longer than this, though fatigue may set in. It is important to note that for this study all interviews were conducted comfortably in one session, regardless of modality. Participants reported that they preferred the telephone interview due to its convenience and the length of the interview did not cause significant concern. Thus telephone interviews have different characteristics compared to face-to-face interviews,

but are largely proven to yield comparable results (Siemiatycki 1979). These results are consistent with the findings of this study.

As with any interview or rating scale that is applied twice within a reasonably tight timeline by the same interviewer, the issue of order effect is a significant one; however, in this study we found none. This may be because a single interviewer performed all of the face-to-face and telephone interviews. This method was chosen to reduce inter-rater biases that may obscure the data. However, this resulted in a lack of blindness of the interviewer to the participant's diagnosis. The first interview was always performed blind, but while the second interview was performed on average a month later, the interviewer was potentially able to recall the diagnosis reached as a result of the first interview. While the interviewer could be expected to forget many of the details of each case over the month or more between interviews, overall impressions and biases are difficult to erase in that time. For example it is possible that the interview that is done second might benefit from the knowledge gained in the first interview. In an effort to reduce these order effects, the interviewer did not compute the algorithm scores for any of the interviews until all interviews had been completed. This minimized the amount of information the interviewer had about the individual's diagnosis prior to any one interview. Furthermore, the interviewer had a full schedule of other ADI-R interviews to complete during that time period and reports that it was difficult to recall and predict responses to items on the ADI-R. While these exigencies may have contributed to better data collection, it is still possible that some bias may have clouded the second interviews. This is why it is significant that counterbalancing of interview modality was practiced, and that no order effects were found in the data.

This study focused on establishing the reliability of the ADI-R by telephone interview for use in research. Future studies should consider inclusion of a control sample of individuals without autism in order to examine the diagnostic specificity of the telephone ADI-R. For research purposes, the ADI-R is often used as a confirmatory diagnostic tool with individuals who have already received, or are strongly suspected of having, an ASD diagnosis. It will be useful to examine whether the ADI-R telephone interview can be reliably and validly used to exclude the ASD diagnosis in research participants for control purposes.

The results presented here indicate that when the ADI-R is used to confirm an existing diagnosis for the purpose of research, the telephone modality is as good as the traditional face to face administration. However, it should be noted that when used for individual scores or when previous diagnoses are not clear, the administration of the ADI-R over the telephone is not indicated. Telephone ADI-R administration is not a substitute for face to face clinical

judgment and these data do not indicate its use for individual assessment or diagnosis on a clinical basis.

Conclusions

This study indicates that, for research purposes, administration of the ADI-R as a confirmatory diagnostic tool can be carried out by trained interviewers over the telephone in place of the traditional face-to-face interview. Given the substantial advantages of the use of telephone over face-to-face interviews, this finding is significant and overdue.

Acknowledgments We extend our sincere appreciation to our research subjects and their extended family members for their enthusiastic support of this study. This work was supported by an OMHF grant (JJAH, principal investigator) and a CIHR Interdisciplinary Health Research Team grant (RT-43820) to the Autism Spectrum Disorders Canadian-American Research Consortium (ASD-CARC: www.autismresearch.ca) (JJAH, principal investigator). JW-K is a postdoctoral trainee with the CIHR/Autism Speaks Autism Spectrum Interdisciplinary REsearch (ASPIRE) Strategic Training in Health Research Program (PI: JJAH) (www.AutismTraining.ca).

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Barriball, K. L., Christian, S. L., While, A. E., et al. (1996). The telephone survey method: A discussion paper. *Journal of Advanced Nursing*, 24(1), 115–121.
- Bauman, L. J. (1993). Collecting data by telephone interviewing. *Journal of Developmental and Behavioral Pediatrics*, 14(4), 256–257.
- Burnard, P. (1994). The telephone interview as a data collection method. *Nurse Education Today*, 14, 67–72.
- Carr, E. C. J., & Worth, A. (2001). The use of the telephone interview for research. *NT Research*, 6(1), 511–524.
- Corey, C. R., & Freeman, H. E. (1990). Use of telephone interviewing in health care research. *HSR: Health Services Research*, 25(1), 129–144.
- Lavrakas, P. J. (1987). *Telephone survey methods: Sampling, selection and supervision*. Thousand Oaks, CA: Sage.
- Law, M. (1997). A telephone survey of day surgery eye patients. *Journal of Advanced Nursing*, 25(2), 355–363.
- Le Couteur, A., Lord, C., & Rutter, M. (2009). Autism diagnostic interview—revised (ADI-R). University of Michigan Autism and Communication Disorders Centre. Retrieved from: http://www.umaccweb.com/diagnostic_tools/adiinfo.html Retrieved Oct 18, 2009.
- Le Couteur, A., Rutter, M., Lord, C., & Rios, P. (1989). Autism diagnostic interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders*, 19(3), 363–387.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview—revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685.
- Mahoney, W. J., Szatmari, P., MacLean, J. E., Bryson, S. E., Bartolucci, G., Walter, S. D., et al. (1998). Reliability and

- accuracy of differentiating pervasive developmental disorder subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37(3), 278–285.
- Marcus, A. C., & Crane, L. A. (1986). Telephone surveys in public health research. *Medical Care*, 24(2), 97–112.
- Musselwhite, K., Cuff, L., McGregor, L., & King, K. M. (2007). The telephone interview is an effective method of data collection in clinical nursing research: A discussion paper. *International Journal of Nursing Studies*, 44(6), 1064–1070.
- Oppenheim, A. N. (1992). *Questionnaire design and attitude measurement*. New York: Basic Books.
- Polit, D. F., & Hungler, B. P. (1991). *Nursing research: Principles and methods* (4th ed.). Philadelphia: Lippincott.
- Robson, C. (1993). *Real world research*. Oxford: Blackwell.
- Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R: The autism diagnostic interview—revised*. Los Angeles, CA: Western Psychological Services.
- Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health*, 69(3), 238–245.
- Tausig, J. E., & Freeman, E. W. (1988). The next best thing to being there: Conducting the clinical research interview by telephone. *American Journal of Orthopsychiatry*, 58(3), 418–427.
- Vrancic, D., Nanclares, V., Soares, D., Kulesz, A., Mordzinski, C., Plebst, C., et al. (2002). Sensitivity and specificity of the autism diagnostic inventory-telephone screening in Spanish. *Journal of Autism and Developmental Disorders*, 32(4), 313–320.
- Waterman, H., Leatherbarrow, B., & Slater, R. (1999). Postoperative pain, nausea and vomiting: Qualitative perspectives from telephone interviews. *Journal of Advanced Nursing*, 29(3), 690–696.
- Wilson, K., Roe, B., & Wright, L. (1998). Telephone or face-to-face interviews? A decision made on the basis of a pilot study. *International Journal of Nursing Studies*, 35, 314–321.
- World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva: WHO.
- Worth, A., & Tierney, A. J. (1993). Conducting research interviews with elderly people by telephone. *Journal of Advanced Nursing*, 18, 1077–1084.