Published in final edited form as:

J Autism Dev Disord. 2014 September ; 44(9): 2375–2380. doi:10.1007/s10803-014-2099-5.

Brief Report: Can Metrics of Reporting Bias Enhance Early Autism Screening Measures?

Cora M. Taylor, Alison Vehorn, Hylan Noble, Amy S. Weitlauf, and Zachary E. Warren

Vanderbilt Kennedy Center, Treatment and Research in Autism Spectrum Disorders, Vanderbilt University, Nashville, TN, USA

Cora M. Taylor: cmtaylor1@geisinger.edu

Abstract

The goal of the current study was to develop and pilot the utility of two simple internal response bias metrics, over-reporting and under-reporting, in terms of additive clinical value within common screening practices for early detection of autism spectrum disorder risk. Participants were caregivers and children under 36 months of age ($n = 145$) participating in first-time diagnostic appointments across our clinical research center due to developmental concerns. Caregivers were asked to complete the Modified Checklist for Autism in Toddlers (MCHAT) as well as a questionnaire embedding six response bias indicator questions. These questions were items that in previous clinical studies had been endorsed by an overwhelming majority of parents within clinically identified populations. Results indicated that removal of self-reports indicative of potential response bias dramatically reduced both false positives and false negatives on the MCHAT within this sample. This suggests that future work developing internal metrics of response bias may be promising in addressing limits of current screening measures and practices.

Keywords

Autism; Screening; Early identification; Internal metrics

Introduction

With an estimated prevalence of one in 88 (CDC 2012) the accurate and effective identification of young children with autism spectrum disorder (ASD) represents a pressing public health and clinical care issue. An increasing body of literature supports substantial gains in cognitive and adaptive functioning for groups of young children with ASD who receive early intensive autism-specific intervention services (Dawson et al. 2010; Warren et al. 2011). However, epidemiological data suggest the average age of diagnosis (CDC 2012)

© Springer Science+Business Media New York 2014

Correspondence to: Cora M. Taylor, cmtaylor1@geisinger.edu.

Present Address: C. M. Taylor, Autism & Developmental Medicine Institute, Geisinger Health System, 120 Hamm Drive, Suite 2A, Lewisburg, PA 17837, USA

Present Address: H. Noble, University of Alabama, Tuscaloosa, AL, USA

A. S. Weitlauf, Department of Pediatrics, Vanderbilt University, TN, USA

Z. E. Warren, Departments of Pediatrics and Psychiatry, Vanderbilt University, Nashville, TN, USA

is still far beyond the age at which ASD may be accurately diagnosed (Corsello et al. 2013). To address the gap between concern and diagnosis, and to take advantage of the potential impact of early intervention, several professional consensus panels have issued practice parameters endorsing early ASD screening. Most recently, the American Academy of Pediatrics endorsed formal screening for ASD at 18 and 24 months (Johnson and Myers 2007).

There are presently several empirically supported screening measures available for assisting in differentiating ASD from typical development in toddlers in this explicit age range with extensive reviews of the psychometric properties of these instruments available in the literature (Barton et al. 2012). The Modified Checklist for Autism in Toddlers (Robins and Fein 1999) and the Infant Toddler Checklist (Wetherby and Prizant 2002) have emerged as the screening instruments that have shown the most promise as ASD-specific screeners in community samples to date (Chlebowski et al. 2013; Miller et al. 2011; Zwiagenbaum 2011). The MCHAT is a 23-item parent report checklist designed to identify symptoms of ASD in children 16–30 months. Initial concerns about high false positive rates based on the checklist alone prompted the authors of the instrument to utilize a companion interview (MCHAT–FUP). This follow-up interview, in combination with the checklist, dramatically enhances positive predictive value (Pandey et al. 2008). The ITC is a 24-item parent checklist designed to detect communication delays children as a component of the larger Communication and Symbolic Behavior Scales—Developmental Profile (Wetherby et al. 2008). Given this context of a broad communication screen, relatively few studies have used the ITC as a standalone ASD screener, with a majority of the literature suggesting providers rely on the MCHAT as an ASD screening tool in standard community practice (Swanson et al. 2013).

In a recent study examining the use of the MCHAT and ITC in community pediatric practice, Miller et al. (2011) demonstrated that both instruments can identify children with ASD earlier than parental/provider concern or basic developmental surveillance alone. However, both required the use of follow-up interviews (completed by trained research staff) to reduce a substantial false positive rate. Even with these procedures, not all children were identified accurately. Importantly, the authors noted that without follow-up interview confirmation, over 100 possibly “unnecessary evaluations” might have been scheduled from the total sample of 796. In another recent work representing the largest community screening protocol to date in the United States, Chlebowski et al. (2013) screened 19,989 toddlers at pediatric well-child visits. Again, screen-positive children received the MCHAT–FUP and those continuing to receive positive screens after completion of the MCHAT–FUP received a diagnostic evaluation. Results suggested that roughly half of the children (54 %) who screened positive via this enhanced procedure presented with ASD, while many (46 %) had other differing or no concerns.

Cumulatively, the best existing data regarding the most widely used screening instruments (i.e., MCHAT and ITC) suggest that standardized screening instruments can identify many children with ASD at early ages. There is also evidence that use of such instruments, at least the self-report questionnaire component, is feasible across well-child settings. Although in many respects this standardized screening method represents a tremendous advance

regarding the early identification of young children with ASD, the same body of research that supports these screeners also documents several potent psychometric, pragmatic, and resource issues that limit current application. First, screening questionnaires alone will not identify many children with ASD (i.e., under-identification/false negatives), with the best available data still incapable of providing population estimates of these errors (see Zwiagenbaum 2011). Second, using questionnaires alone, without utilizing *uncommon research procedures* to verify findings, results in substantial over-identification rates with perhaps more than one in ten children identified at-risk for ASD (Miller et al. 2011). This is particularly problematic as is shown in recent work, which documents extremely low rates of MCHAT-FUP use or other verification procedures in community pediatric settings (Swanson et al. 2013). Third, even when using verification procedures, screening initially identifies large numbers of children *without* ASD as at-risk (Chlebowski et al. 2013; Miller et al. 2011). These results highlight the fact that simple questionnaire-based screening alone, without a method for verifying the nature of concerns, may be extremely problematic in widespread practice.

In order to improve the accuracy and validity of screening measures for complex neurobehavioral conditions, other self-report and parent-report tools have embedded metrics of response characteristics and bias. Such metrics are often included to account for response characteristics that may influence the scores obtained by the instruments themselves, such as over-reporting, under-reporting, or inconsistently reporting symptoms. The Minnesota Multiphasic Personality Inventory, Second Edition [MMPI-2] (Butcher et al. 2001) is thought of as the exemplar self-report instrument for understanding reporting characteristics in that it contains not only simple methods for documenting valid profiles, but also methods for adjusting scores and interpretation based on response patterns. Another widely-used instrument for indexing challenging behavior in children, the Behavior Assessment System for Children, Second Edition [BASC-2] (Reynolds and Kamphaus 2006), has adopted a similar approach to documenting parental response characteristics that may affect interpretation of scores. Such an approach of developing internal metrics for assessing validity of parental reports of concern may be extremely valuable to consider for application within ASD screening processes, given that research suggests that (1) common measures attempting to index ASD specific symptoms are often prone to elevations in scores due to non-ASD behavioral concerns and potentially parenting stress (Hus et al. 2013; Warren et al. 2011) and (2) many providers administering ASD screeners are not utilizing recommended structured follow-up procedures (Swanson et al. 2013).

The goal of the current study was to develop and pilot the utility of two simple internal response bias metrics, over-reporting and under-reporting, in terms of additive clinical value within common screening practice (i.e., MCHAT questionnaire use). We selected items from a general developmental questionnaire endorsed by parents from an identified clinical population at a very low frequency (over-reporting) and very high frequency (under-reporting). We then had parents of children from a new cohort of clinically referred children complete these items along with the MCHAT to assess whether utilizing these response bias items would enhance screening accuracy within this sample. We hypothesized that utilizing

response bias items would help identify both false positive and false negative reports on the MCHAT questionnaire.

Methods

Participants and Design

This sample was drawn from a university-based clinical research center for autism. Caregivers ($N = 145$) of children under 36 months of age (mean age = 2.34 years; $SD = 0.45$) who were referred for a psychological evaluation were asked to fill out both the MCHAT questionnaire as well as a developmental questionnaire with embedded response bias items prior to their child's scheduled evaluation. We explicitly did not utilize the MCHAT follow-up interview as part of this process. We simply scored the MCHAT as pass/fail based on the suggested scoring algorithm of total score three or greater or critical item score of two or greater on the questionnaire (Robins et al. 2001). As part of the evaluation process, children received a clinical best estimate diagnosis from participating psychological providers based on an assessment that involved a clinical interview, cognitive assessment, adaptive behavior assessment, and a research reliable administration of the Autism Diagnostic Observation Schedule (Lord et al. 2001; see Table 1 for summary of participant characteristics).

Measure Development

In order to develop our pilot response bias questions we adopted a simple strategy for identifying common and uncommon response patterns within a clinically identified sample. Specifically, we pulled the high and low response items from a general developmental survey administered across clinical research programs at our university autism center. Some 235 questionnaires were available for children under 36 months of age who were identified with ASD (94 %) or an associated neuro developmental disorder (6 %). Although these surveys were not formally scored, they contained information about basic developmental skills and abilities that could be quantified and scored (i.e., several yes/no and Likert items), in addition to qualitative responses. We operationalized and selected potential response bias questions as dichotomous items (i.e., yes/no; true/false) that were most commonly endorsed by parents. Specifically, potential over-reporting response bias questions were endorsed by less than 10 % of the development sample (range 9.1–4.8 %) and under-reporting items were endorsed by over 90 % of the development sample (range 90.8–93.3 %). This selection method and procedure allowed us to extract items that would be endorsed or not endorsed at very high frequencies for the identified patient population (i.e., skills that were likely almost always within or below developmental expectations for a referred clinical population during this specific developmental window).

We embedded these response bias questions within an additional short general developmental questionnaire in a simple attempt to ensure items were not construed as standalone validity checks by parents. This final measure asked 14-general questions about development (e.g., Is your child using words to communicate?/Is your child using phrases to communicate?/How often do you see your child pretending while playing?/Are you concerned about your child's motor development?) as well as six identified reporting bias

indicator true–false questions (*Over-reporting*: “My child can roll or throw a ball,” FALSE; “My child will laugh and smile when happy,” FALSE; “My child frequently holds his breath when upset,” TRUE; *Under-reporting*: “My child understands everything we say,” TRUE; “My child frequently makes up complex stories when pretending,” TRUE; “My child will often imitate complex actions with dolls and action figures,” TRUE).

Analytic Strategy

We operationalized validity concern for over-reporting or under-reporting as endorsing any one item from these particular indices. Subsequently, we examined whether corresponding reductions in false positives (i.e., over-reporting of symptomology) or false negatives (i.e., under-reporting of symptomology) on the MCHAT were seen when these scores were incorporated into identification processes.

Results

MCHAT Questionnaire Use

Of the 145 children participating in the current study, 86 were diagnosed with ASD, 41 received a developmental or behavioral diagnosis other than ASD, and 18 children were not diagnosed with a developmental or behavioral disorder. When examining the properties of the MCHAT questionnaire alone across the entire sample, 100 children scored above cut-off (failed) on the instrument. Of these, 74 children were given an ASD diagnosis, and 26 received a developmental diagnosis apart from ASD (false positives). Some 45 children in our sample scored below cut-off (passed) on the MCHAT, with 18 of these children carrying no developmental diagnosis, 15 receiving a developmental diagnosis apart from ASD, and 12 receiving a diagnosis of ASD (false negatives). In total, 107 of 145 (74 %) participating children were accurately classified with this instrument with 74 of 86 (86 %) of children with ASD accurately identified as at-risk.

Response Bias Analysis

False Negatives—Twelve children with ASD (14 % of ASD sample) passed the MCHAT questionnaire. Within this group, half ($n = 6$) of the parents had endorsed one or more of the identified under-reporting items (Mean = 2.00; SD = 0.63). As such, withdrawing these profiles based on the response bias indicator resulted in a 50 % reduction of the number of false negatives of MCHAT questionnaire use for individuals with ASD. When examining the entire sample of children with ASD, false negatives represent 7.5 % of the ASD sample after correction, a decrease of 46 % (See Table 2).

False Positives—Twenty-six children with a developmental or behavioral diagnosis other than ASD (44 % of the sample) failed the MCHAT questionnaire. Within this group, nine parents had endorsed one of the identified over-reporting items. No parent endorsed more than one of the over-reporting validity items. Withdrawing these profiles based on the validity flag results in a 34 % reduction in the number of false positives. Looking at the entire sample of children with a diagnosis other than ASD, false positives represent 34 % of the non-ASD sample after correction, a decrease of 22.7 % (See Table 2).

Discussion

The current study is consistent with a growing body of literature suggesting that the use of a simple standardized screening tool for ASD at young ages will be helpful in accurately identifying many young children with ASD. However, our findings also are consistent with this same literature base, suggesting that simple use of a common ASD screening questionnaire in absence of a validation procedure will result in both (1) over-identification of many children who do not in fact have ASD and (2) a failure to identify risk in some children with ASD. The MCHAT and other screening instruments have documented reductions in over-identification when follow-up interviews are utilized in practice. Given that these follow-up interviews may not necessarily be utilized in wide-scale community practice (Swanson et al. 2013), the current study evaluated whether simple response bias metrics might prove an additional strategy for improving the accuracy of screening instruments. Importantly, results documented that flagging self-reports on the basis of items endorsed with high or low frequency within an identified clinical sample dramatically reduces both false positives (34 % reduction) and false negatives (50 % reduction). These preliminary findings suggest that there may be great value in explicitly focusing on developing metrics of internal validity and response bias that may help providers more specifically screen and triage families for additional evaluations across resource-strained environments.

It is extremely important to acknowledge that the current investigation represents only an initial pilot investigation evaluating the possibility and value of further work toward developing internal metrics of response bias that enhance psychometric validity of self-report screening measures for ASD. We acknowledge that the approach utilized here, examining low and high frequency items derived from a defined clinical set, represents only a preliminary study in that regard. As such, there are extreme limitations of the current approach in terms of identifying a final item set or scale that is ready to function independently within current community practice with current screening tools. Specifically, we have no data on how such a set or scale would function within a low-risk general population sample, in that we tested this response bias questions in a sample of children already referred for evaluation. Studying such metrics in such a population would be critical to determine ultimate benefit of how such scales could reduce identification errors. We also have no data to suggest that this validation procedure would provide an enhancement in reducing false positives relative to the already defined MCHAT-FUP interview. Further, while many children often identified at-risk who are not diagnosed with ASD are discussed as false positives, often these children have other developmental concerns that may warrant attention. Thus, labeling all diagnostic cases as false positives in this regard as errors may not be appropriate. Furthermore, this data was pulled from a database with limited demographic information on the children and families represented by the sample. Without this information, we cannot fully generalize that these over- and under-reporting items would function accurately across different ethnic backgrounds, race, socio-economic status, etc. Despite these limitations, our study documented that errors of under-identification, wherein current screening measures have no identified validity check, might be reduced by utilizing scales or items that flag such response tendencies on screeners themselves. In

addition, knowing that many providers and systems are not utilizing validation interviews after positive questionnaire screens across care systems, our findings suggest that further developing and utilizing items or scales that flag reporting bias in this direction have value in reducing potentially inappropriate referrals for some children and practice settings.

Issues of both over-identification and under-identification of screening processes are extremely problematic when pushing for early accurate ASD diagnosis on a population level. Although our field has attempted to implement screening methods that optimize risk identification, the steps subsequent to screening in community settings are much less clear and often problematic for clinicians, families, and systems of care alike. Initiation of population screening with trends toward over-identification without additional resources for subsequent care can unfortunately result in more children being referred for a very limited number of expert diagnostic assessment resources. Further, the paucity of data and methods for understanding the clinical characteristics of screen negative children is also challenging. This suggests the need for additional methods for increasing the accuracy of ASD screening instruments across resource limited environments. Our current findings suggest that future research developing internal metrics of response bias may be one such method for increasing the accuracy of autism screening measures.

Acknowledgments

This work was supported in part by the Vanderbilt Kennedy Center and Vanderbilt University Medical Center. This includes core support from CTSA award No. UL1TR000445 from the National Center for Advancing Translational Sciences and the Eunice Kennedy Shriver National Institute of Child Health and Human Development [R01 HD057284-02 and R01 HD039961]. Support was also provided by Health Resources and Service Administration's Maternal Child Health Bureau LEND Grant [T73MC00050]. Portions of this paper were presented at the 2013 International Meeting of Autism Research, San Sebastian, Spain.

References

- Barton ML, Dumont-Mathieu T, Fein D. Screening young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental Disorders*. 2012; 42:1165–1174.10.1007/s10803-011-1343-5 [PubMed: 21842325]
- Butcher, JN.; Graham, JR.; Ben-Porath, YS. *Minnesota multiphasic personality inventory-2*. Minneapolis: University of Minnesota Press; 2001.
- CDC. Prevalence of autism spectrum disorders—Autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *MMWR*. 2012; 61(03):1–19.
- Chlebowski C, Robins DL, Barton ML, Fein D. Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*. 2013;110.1542/peds.2012-1525
- Corsello CM, Akshomoff N, Stahmer AC. Diagnosis of autism spectrum disorders in 2-year-olds: A study of community practice. *Journal of Child Psychology and Psychiatry*. 2013; 54:178–185. [PubMed: 22905987]
- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*. 2010; 125:e17–e23. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19948568>. [PubMed: 19948568]
- Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2013; 54(2):216–224.10.1111/j.1469-7610.2012.02589.x
- Johnson CP, Myers SM. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*. 2007; 120:1183–1215.10.1542/peds.2007-2361 [PubMed: 17967920]

- Lord, C.; Rutter, M.; DiLavore, PC. Autism diagnostic observation schedule. Torrance: Western Psychological Services; 2001.
- Miller JS, Gabrielsen T, Villalobos M, Alleman R, Wahmhoff N, Carbone PS, Segura B. The each child study: Systematic screening for autism spectrum disorders in a pediatric setting. *Pediatrics*. 2011; 127:866–871. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21482605>. [PubMed: 21482605]
- Pandey J, Verbalis A, Robins DL, Boorstein H, Klin A, Babitz T, et al. Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*. 2008; 12:513–535. Retrieved from <http://aut.sagepub.com/cgi/content/abstract/12/5/513>. [PubMed: 18805945]
- Reynolds, CR.; Kamphaus, RW. Behavior assessment system for children. 2nd. Upper Saddle River: Pearson Education; 2006.
- Robins DL, Fein D. Modified Checklist for Autism in Toddlers. 1999
- Swanson AR, Warren ZE, Stone WL, Vehorn AC, Dohrmann E, Humberd Q. The diagnosis of autism in community pediatric settings: Does advanced training facilitate practice change? *Autism*. 2013;10.1177/1362361313481507
- Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-Vanderweele J. A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*. 2011; 127:e1303–e1311. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21464190>. [PubMed: 21464190]
- Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. Validation of the Infant–Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism the international journal of research and practice*. 2008; 12:487–511.10.1177/1362361308094501 [PubMed: 18805944]
- Wetherby, AM.; Prizant, BM. Infant toddler checklist. Baltimore: Paul H. Brookes; 2002.
- Zwiagenbaum, L. Screening, risk, and early identification of autism spectrum disorders. In: Amaral, DG.; Dawson, G.; Geschwind, DH., editors. *Autism spectrum disorders*. New York: Oxford Press; 2011. p. 75-89.

Table 1
Participant characteristics

	Typical development M (SD)	ASD M (SD)	Other diagnosis M (SD)
Age	2.3 (0.59)	2.3 (0.43)	2.4 (0.46)
Mullen composite	104.50 (13.46)	58.47 (15.76)	70.76 (16.47)
Vineland composite	96 (25.46)	68.78 (7.87)	81.61 (36.06)
ADOS composite	3.10 (2.38)	20.09 (5.02)	6.34 (4.89)

Table 2
Incorrect identification by diagnosis

	Typical development	ASD	Other diagnosis
<i>n</i>	18	86	41
Correctly classified	18	74	15
Incorrectly classified	0	12	26
Incorrectly classified after correction	0	6	17
% Decrease of incorrect classifications	n/a	50 %	34 %
% Decrease relative to entire diagnostic group	n/a	46 %	22.7 %

For children with ASD, incorrect classification represents a false negative. In children with typical development or other diagnoses, incorrect classification represents a false positive