ORIGINAL PAPER

Evaluation of a Records-Review Surveillance System Used to Determine the Prevalence of Autism Spectrum Disorders

**Rachel Nonkin Avchen · Lisa D. Wiggins ·
Owen Devine · Kim Van Naarden Braun ·
Catherine Rice · Nancy C. Hobson · Diana Schendel ·
Marshalyn Yeargin-Allsopp**

Abstract We conducted the first study that estimates the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of a population-based autism spectrum disorders (ASD) surveillance system developed at the Centers for Disease Control and Prevention. The system employs a records-review methodology that yields ASD classification (case versus non-ASD case) and was compared with classification based on clinical examination. The study enrolled 177 children. Estimated specificity (0.96, [$CI_{.95}$ = 0.94, 0.99]), PPV (0.79 [$CI_{.95}$ = 0.66, 0.93]), and NPV (0.91 [$CI_{.95}$ = 0.87, 0.96]) were high. Sensitivity was lower (0.60 [$CI_{.95}$ = 0.45, 0.75]). Given diagnostic heterogeneity, and the broad array of ASD in the population, identifying children with ASD is challenging. Records-based surveillance yields a population-based estimate of ASD that is likely conservative.

Keywords Autism · Surveillance · Sensitivity · Specificity · Prevalence

R. Nonkin Avchen (✉) · L. D. Wiggins · O. Devine ·
K. Van Naarden Braun · C. Rice · D. Schendel ·
M. Yeargin-Allsopp
Developmental Disabilities Branch, Centers for Disease Control
and Prevention, 1600 Clifton Road, MS E-92, Atlanta, GA
30333, USA
e-mail: ravchen@cdc.gov

N. C. Hobson
Research Triangle Institute, Atlanta, GA, USA

Introduction

Children with autism spectrum disorders (ASD) are characterized by deficits in social, communication, and behavior skills, but the level of deficit in each of these domains differs for each child (American Psychiatric Association 1994). Because of the heterogeneous nature of ASD, and the fact that they are behaviorally defined, surveillance and epidemiological investigations can be quite challenging. As recently as the early 1990s, ASD prevalence was considered rare, at only 0.4–0.5 per 1,000 children (Gillberg et al. 1991; Fombonne 1996; Rutter 2005). However, recent population-based studies report ASD prevalence from 4.2 to 12.1 per 1,000 among children 8 years of age in the United States, with an average prevalence across 11 US communities of 9.0 per 1,000 children (Centers for Disease Control and Prevention 2009). Beyond the United States, a similar range has been estimated (Baird et al. 2006; Kadesjo et al. 1999; Kawamura et al. 2008; Posserud et al. 2006).

Variation in ASD surveillance methods may provide an explanation for the reported range of prevalence estimates. Four surveillance methods have typically been implemented to identify ASD cases. Each has its limitations. While *clinical screening and comprehensive clinical evaluation* are viewed as the "gold standard" for detecting and diagnosing ASD, the population validity of such an approach is dependent on unbiased participation, sample size, location, and infrastructure. There are also potential resource constraints given the need for clinical evaluations, which require qualified professionals and parent/child attendance. *Surveys* that rely on parental report of ASD diagnosis do not verify clinical diagnosis and results from telephone surveys are biased by participants with landline telephones (Kogan et al. 2009).

Survey sample size is also generally insufficient to provide estimates at the state or local population-level as evidenced by the National Survey for Children's Health, one of the largest national surveys of children's health that strives to enroll approximately 1,000 children per state but is deemed insufficient to develop state-based estimates for ASD. Further, limited or no data are collected on ASD subtype, co-occurring conditions (e.g., intellectual disability), and specific phenotypic characteristics of children identified as having ASD (Schieve et al. 2006). Review of *service registries and administrative databases* for children diagnosed with or served for ASD capture only those children receiving ASD services or with particular service codes (Croen et al. 2002). A *multi-source records-review of health and education records* is laborious but might increase the accuracy of prevalence estimates. Applying a standard case definition based on review of behaviors described in a child's past developmental evaluations, and thus expanding classification beyond only children with a previously documented diagnosis, might further increase sensitivity. This method is also more cost effective than a similar population-based clinical screening and evaluation approach.

Providing accurate prevalence estimates is a critical step toward understanding the impact of ASD on the population and will quantify the need for public health action. Data generated from prevalence reports continue to be used to guide research and policy decisions and inform hypotheses for epidemiologic research. Further, ongoing and uniform surveillance methods for the broad array of ASD facilitate in-depth analyses on trends in ASD prevalence and on characteristics of children identified by the surveillance system. The Autism and Developmental Disabilities Monitoring (ADDM) Network uses methods adapted from the Centers for Disease Control and Prevention's (CDC's) Metropolitan Atlanta Developmental Disabilities Surveillance Program (MAD-DSP; Rice et al. 2007; Van Naarden Braun et al. 2007). ADDM is an ongoing, active surveillance system that uses records-review methods in the United States. This system has not been rigorously evaluated by comparing case status obtained by the surveillance approach with those based on clinical evaluation. Thus, the primary goal of our study was to provide the first estimates of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of ASD case definitions derived from MADDSP. We also sought to examine (a) how our clinical case definition, which was based on results from the Autism Diagnostic Interview—Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS), influenced sensitivity, specificity, PPV, and NPV estimates, and (b) whether the proportion of children misclassified by the records-review method (who were found

to have an ASD during the clinical evaluation) differed from those who did or did not have an ASD diagnosis documented by a community professional.

Methods

Study Design and Participant Ascertainment

In this study, we used a cohort design with a two-stage sampling process, herein referred to as Phase I and Phase II. The Phase I outcome was defined by results of MAD-DSP records-review case status methods, while the Phase II outcome was determined from attendance at a comprehensive clinical evaluation.

Phase I

MADDSP case ascertainment for ASD surveillance was based on systematic review of health and education records of children born in 1997 who resided in the catchment area of one public school district in metropolitan Atlanta in 2005. The school district is part of the most populous Georgia County and is included in the core, five-county MADDSP surveillance area (Boyle et al. 1996) and was chosen because, at the time of study design, it was comparable to the larger MADDSP surveillance area in gender and race/ethnicity distributions. All records of children with suspected or diagnosed ASD and related neuro-developmental disorders, who met residency and birth-year criteria, were reviewed at the time the child was 8 years old. For example, school records of children receiving special education services for autism, intellectual disability, learning disability, or other health impairments were reviewed; likewise, any files from collaborating healthcare settings for children with a broad range of diagnostic codes (e.g., from 299.00 for ASD to 759.00 for other congenital anomalies such as Tuberous sclerosis or Fragile X) were examined (CDC, ADDM Network 2007). Surveillance personnel conducted a review of these records for an ASD trigger, which is one of many social impairments related to ASD (e.g., limited eye contact or no interest in other children). For those records wherein an ASD trigger was identified, verbatim descriptions of developmental, behavioral, psychological, and relevant medical data and diagnoses were abstracted. Abstractions from all pertinent education and/or medical records for each child were compiled into a single record, which represented that child's surveillance data.

All information collected from surveillance records was reviewed by trained clinicians who applied a standardized coding scheme (CDC, ADDM Network 2007) to abstracted records irrespective of any documented diagnosis that was

consistent with the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) criteria for Autistic Disorder, PDD-NOS, or Asperger's disorder. To avoid classification bias, Phase I records-review clinicians were independent of the Phase II clinical evaluation clinician. Children were classified as an ASD surveillance case if results of the systematic classification of behaviors by qualified reviewers agreed with the DSM-IV-TR surveillance algorithm (described in CDC, ADDM Network 2007). If a child was classified as an ASD surveillance case, but the primary clinician reviewer disagreed with the scoring algorithm, a second and independent review of the record was requested. A consensus discussion between the primary and secondary reviewers resolved the final surveillance case status. Initial case status could be changed to non-case status if both reviewers felt that ASD symptoms were better accounted for by another disorder. Secondary reviews were not initiated for children who did not meet ASD surveillance case status during the initial, primary review.

Note that ASD surveillance case status was based on behaviors noted in abstracted records and not solely on the presence of a documented ASD diagnosis. Documented ASD diagnoses were identified from information abstracted from the surveillance records, representing how the children were classified by qualifying community professionals (e.g., psychologists, developmental pediatricians), and independent of their surveillance case status. ASD diagnoses were recorded only if the authoring community professional specifically stated that the child met ASD diagnostic criteria. Hence, a documented ASD diagnosis was considered when determining surveillance case status but was not necessary to meet ASD surveillance criteria; that is, as long as there was sufficient evidence of behaviors that met the standardized coding criteria, a child could be called a surveillance case despite the lack of a formal diagnosis noted in the record.

For purposes of this study, children in Phase I were classified into one of three surveillance strata based on (a) records that were reviewed but not abstracted, because there was no ASD trigger (not abstracted/non-ASD case), (b) records that were reviewed and abstracted but did not meet ASD surveillance case status after final clinician review (abstracted/non-ASD case), and (c) records that were reviewed, abstracted, and met surveillance case status after final clinician review (abstracted/case; see Fig. 1).

Phase II

The second phase of the study involved active tracing and recruitment of the Phase I population to participate in a clinical evaluation. A targeted sampling approach from the three surveillance strata was necessary to ensure feasibility
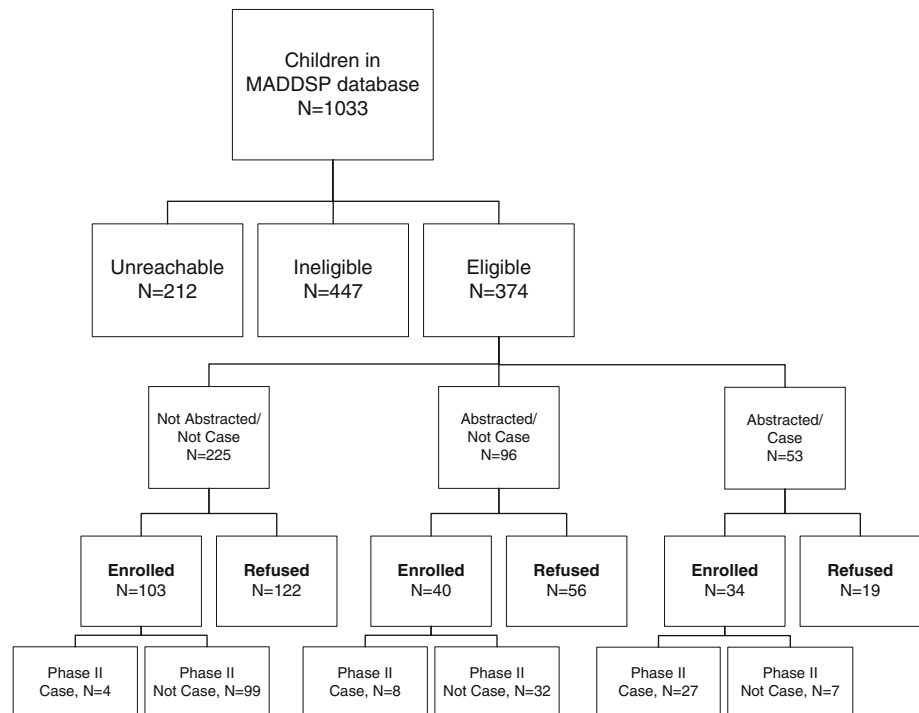
and efficiency of this phase of the study. To address potential verification bias (Zhou 1998), we used a weighted approach outlined by McNamee (2002) wherein sampling from each Phase I strata was determined by the maximum number of clinical evaluations deemed feasible during Phase II and expected ASD prevalence in the Phase I population. We thought 250 evaluations were feasible, given resource and time constraints. The 250 Phase II participants were allocated among Phase I stratum to produce a minimum possible standard error for the sensitivity estimate under the total sample size constraint. Sampling from the Phase I participants was also based on the assumptions that (a) ASD prevalence in the total population of 8-year-olds ranged from 0.3 to 0.8%, (b) 85% of surveillance records reviewed would not contain triggers for abstraction, (c) 5% of surveillance records reviewed would be abstracted, but not classified as a case after clinician review, and (d) 10% of surveillance records reviewed would be abstracted and deemed a surveillance case (estimates were based on MADDSP ASD data for the surveillance year 2000). In addition, we assumed that 1% of the children with no ASD triggers found in surveillance records during Phase I (i.e., not abstracted/non-ASD case) would be classified as a clinical case in Phase II. Rather than assign a single set of values for the probability a child in the abstracted/non-ASD case or abstracted/case strata would meet the Phase II clinical case definition, we considered a range of values for this parameter that were consistent with the aforementioned assumptions.

Under these assumptions, and given a total Phase II sample of 250 participants, the optimum stratum-specific sample sizes were estimated to be 200 (not abstracted/non-ASD case), 34 (abstracted/non-ASD case), and 16 (abstracted/case strata) subjects. We evaluated the impact of departures from the assumptions underlying determination of optimal sample size using simulation methods. This assessment indicated that reasonable departures from the assumptions had minimal effect on the variance of the estimated sensitivity. In addition, we created simulations to assess the potential impact of recruiting fewer participants than constituted the optimal sample. Reduction in sample size of up to 50% did not substantially bias the records-review sensitivity estimates; however, results from 1,000 simulations suggested that 75% of the simulations had a 90% confidence interval that contained the true assumed value for the sensitivity parameter.

Study Population and Recruitment of Study Participants

Research procedures were approved by CDC's institutional review board and met Health and Human Services policy for protection of human research subjects. In an effort to reduce Phase I records-review burden and to implement

Fig. 1 Phase I records-review stratum and Phase II clinical evaluation enrollment status. *MADDSP*, Metropolitan Atlanta Developmental Disabilities Surveillance Program

Phase II in a timely fashion, the five-county metropolitan Atlanta surveillance area was limited to one-county for the purpose of our study. Therefore, instead of the broader MADDSP five-county area, educational records for only one public school district within the selected county were reviewed for participant ascertainment. All health records identified at partner sources for eligible children residing within the public school boundaries of the selected county were also reviewed.

We identified 1,033 children who were considered potential participants (see Fig. 1). Families were ineligible (*n* = 447) to participate in Phase II if they could not communicate fluently in English or if they no longer resided within a 45-mile radius of the defined study area and thus could not attend a clinical evaluation. Those families who could not be contacted for recruitment, either through mailings or phone solicitation, were considered unreachable (*n* = 212). Therefore, the eligible Phase II cohort was limited to 374 children who met four eligibility criteria: (a) born in 1997, (b) parents resided within the limited catchment area during Phase I records-review, (c) parents still resided in the study area or within a 45-mile radius from the boundary lines during Phase II clinical evaluations, and (d) parents could communicate fluently in English.

Families of all 374 eligible children were contacted to participate in the Phase II clinical evaluation. Our initial goal was to enroll the optimum number of subjects from each Phase I stratum and to discontinue recruitment once the target for that particular group was met. Originally, we

recruited for Phase II by mailing an introductory letter with a prepaid response card to the families with traced residential addresses. Response cards returned with affirmative study interest and cards not received 2 weeks post-introductory mailings were followed by recruitment phone calls. When only 8% of response cards were returned, we deployed initial recruitment via phone contact, and mailed letters only upon request or if an active phone number was not identified. Additionally, to maximize recruitment numbers, any family contacted and willing to participate was enrolled.

Phase II Clinical Measures

An experienced psychologist blinded to Phase I records-review case definitions collected Phase II clinical evaluation data. Data collected on families who agreed to participate in Phase II of the study included cognitive, adaptive, and ASD screening and diagnostic measures comprising the Differential Ability Scales (DAS; Elliot 1990), Vineland Adaptive Behavior Scales-II (Sparrow et al. 1984), the Social Communication Questionnaire (SCQ; Berument et al. 1999), ADI-R (Lord et al. 1994; Rutter et al. 2003), and the ADOS (Lord et al. 1989; Lord et al. 2002). The ADI-R and ADOS are two of the most commonly used instruments to identify children with ASD in clinical and research settings, and these were administered to establish Phase II ASD clinical case status. The ADI-R is a semi-structured parental interview that provides a historical account of social, communicative, and

behavioral symptoms associated with autism, as well as developmental differences before 3 years of age. The ADOS is a child-observation and measurement tool that attempts to elicit social interaction and communication through structured play activities; children are classified as having ASD or autistic disorder for the ADOS scoring algorithm. Utility of both instruments is supported by acceptable psychometric properties, as well as recommended practice guidelines (Johnson et al. 2007; Filipek et al. 2000). The psychologist who administered the clinical evaluations was research reliable, according to University of Michigan Autism and Communication Disorder Center (UMACC) standards on the ASD diagnostic measures, and maintained acceptable reliability via quarterly quality control checks with a second, research-reliable clinician.

Phase I and Phase II Case Definitions

For Phases I and II, we identified each participant as an ASD case or non-ASD case. Research suggests the ADOS and ADI-R provide better sensitivity and specificity when used in conjunction rather than when used alone (Risi et al. 2006). Therefore, a Phase II ASD case was defined as a child who met both ADOS ASD criteria and ADI-R autism criteria for the diagnostic algorithms. To remain consistent with MADDSP methods, which assume a lifetime prevalence of ASD (even if not detected before 3 years of age), the abnormality of development domain (i.e., the domain that measures developmental delay before 3 years of age) was excluded from the ADI-R. A Phase II non-ASD case was defined as one in which the child did not meet both ADOS ASD criteria and ADI-R autism criteria for the diagnostic algorithms (Fig. 2).

Discordant results between ADOS and ADI-R were resolved by relaxing certain portions of the ADOS and ADI-R diagnostic algorithms. Specifically, discordance between the two instruments was resolved if a child: (a) met ADOS algorithm criteria for ASD, met ADI-R social criteria, and was within 2 points of meeting ADI-R communication criteria, (b) met ADOS algorithm criteria for ASD, met ADI-R communication criteria, and was within 2 points of meeting ADI-R social criteria, or (c) met ADOS social criteria for ASD, met ADI-R social criteria, and was within 1 point of meeting ADI-R behavioral criteria. Children whose discordant results were resolved with any of these criteria were considered Phase II cases, whereas children whose discordant results were not resolved by any of these criteria were considered Phase II non-ASD cases. Although these criteria have not been extensively studied and replicated in published reports, they were recommended by other researchers who found acceptable
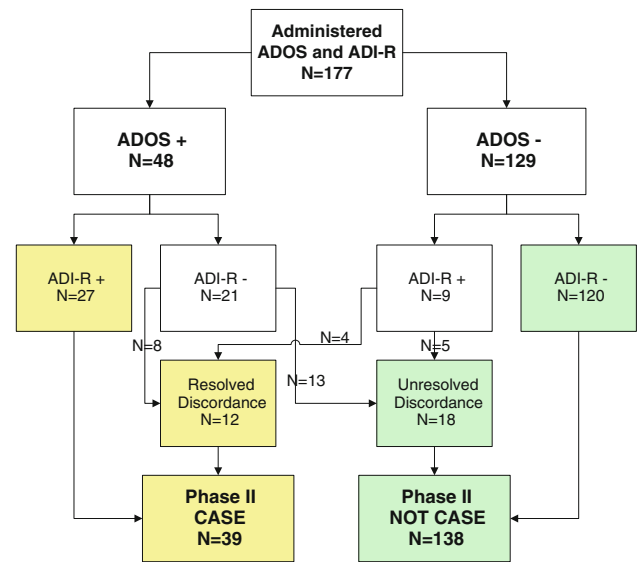

Fig. 2 Phase II case definition. *ADOS*, Autism Diagnostic Observation Schedule; *ADI-R*, Autism Diagnostic Interview—Revised

sensitivity and specificity estimates when discordance rules are applied (Risi et al. 2006).

Statistical Analyses

The primary statistical analysis focused on estimating the sensitivity, specificity, PPV, and NPV of ASD case definitions from Phase I records-review, using results from the Phase II clinical evaluation as the gold standard. Estimates and standard errors for these parameters were derived using methods described by McNamee (2002) to address potential biases associated with differential sampling from various Phase I outcome strata (i.e., not abstracted/non-ASD case, abstracted/non-ASD case, and abstracted/case). In particular, the estimated sensitivity is defined as the proportion of clinical cases correctly classified as ASD in MADDSP records reviews weighted to reflect the sampling probabilities in each of the Phase I outcome stratum. Similarly, specificity was estimated as a weighted combination of Phase II non-ASD cases also classified as Phase I non-ASD cases, with weights again reflecting Phase I sampling fractions. To illustrate the potential impact of the surveillance parameters on the ASD prevalence estimates, we derived an adjustment factor to reflect differences between the estimated prevalence in Phase I and Phase II. The adjustment and an associated 95% CI were estimated using the ratio of ASD prevalence from Phase II (McNamee 2002) to the prevalence derived from Phase I.

In a secondary analysis, we examined the influence of defining a Phase II clinical case as only those children who met ADOS and ADI-R diagnostic criteria, without the benefit of resolving discrepancies between the two

instruments, by estimating sensitivity, specificity, PPV, and NPV given this narrower case definition. We also examined the influence of a documented ASD diagnosis during Phase I records-review among children defined as having an ASD case during the Phase II clinical evaluation, using a χ^2 analysis.

Results

Descriptive Analyses

Of the 374 eligible families, nearly half (177; response rate 47%) agreed to participate and were enrolled in Phase II of the study (see Fig. 1). The families who agreed to participate in Phase II came from the following Phase I records review stratum: 103 (46% of the eligible sample) from the not abstracted/non-ASD case group, 40 (42%) from the abstracted/non-ASD case group, and 34 (64%) from the abstracted/case group (see Fig. 1). Thirty-nine children were classified as ASD cases based on results of the Phase II clinical evaluation (see Fig. 2). The mean age of Phase II participants was 9.4 years; 73% were male and 27% were female. The ethnic/racial distribution of the Phase II sample was 50% non-Hispanic white, 38% non-Hispanic black, and 11% other. The mean cognitive standard score was 89 (range = 25–147); 28% of the sample had intellectual disability (defined as a standard score ≤ 70 points on the DAS).

There was no difference between surveillance stratum by sex or race. However, there were differences in surveillance stratum based on performance on cognitive, adaptive, and ASD measures. Among children with intellectual disability, 42% were not abstracted/non-ASD cases, 29% were abstracted/non-ASD cases, and 29% were abstracted/cases, χ^2 (2, $N = 177$) = 6.43, $p < .04$. Likewise, on the adaptive measure, 15% of children in the not abstracted/non-ASD cases demonstrated below-average performance, as well as, 35% of abstracted/non-ASD cases and 44% of abstracted/cases, χ^2 (2, $N = 177$) = 14.80, $p < .01$. Similar results were found for the ASD diagnostic measures, in that a higher proportion of the abstracted/case group met ASD diagnostic criteria for both the ADOS, χ^2 (2, $N = 177$) = 79.9, $p < .01$, and ADI-R, χ^2 (2, $N = 177$) = 70.2, $p < .01$).

Sensitivity, Specificity, PPV, and NPV

Estimates of specificity (0.96, [CI$_{.95}$ = 0.94, 0.99]), PPV (0.79 [CI$_{.95}$ = 0.66, 0.93]), and NPV (0.91 [CI$_{.95}$ = 0.87, 0.96]) were fairly high, while estimated sensitivity (0.60 [CI$_{.95}$ = 0.45, 0.75]) was lower (see Table 1). Nineteen children were misclassified by Phase I records-review methods based on the results of Phase II clinical evaluation case criteria. Of those misclassified 68% were male, 63% non-Hispanic white, 26% non-Hispanic black, 47% had a cognitive score < 70, and 37% met case definition through the discordance rules (see Table 2). Gender and race proportions were aligned with the overall sample, but intellectual impairment was higher in the misclassified group (47% in the misclassified group had cognitive score < 70, compared with 28% in the overall sample). Further, 58% of

Table 1 Weighted sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for Phase I records-review based on Phase II clinical evaluation

	Parameter	CI$_{.95}$	Phase I records-review stratum			
			Not abstracted ($n = 103$)	Abstracted/not case ($n = 40$)	Abstracted/case ($n = 34$)	Total ($n = 177$)
Phase II ASD cases[a]			4	8	27	39
Sensitivity	0.60	0.45, 0.75				
Specificity	0.96	0.91, 1.00				
PPV	0.79	0.66, 0.93				
NPV	0.91	0.87, 0.96				
Phase II ASD cases[b]			2	3	22	27
Sensitivity	0.75	0.57, 0.92				
Specificity	0.94	0.90, 0.99				
PPV	0.65	0.49, 0.81				
NPV	0.96	0.93, 0.99				

[a] Including those classified using ADOS/ADI-R discordance rules; discordance rules are defined in section labeled "Phase I and Phase II case definitions"

[b] Excluding those classified using ADOS/ADI-R discordance rules; discordance rules are defined in section labeled "Phase I and Phase II case definitions"

Table 2 Characteristics of case children that were misclassified by Phase I records-review

Child	Phase I, records-based status	Phase II, clinical evaluation status	Sex	Race or ethnicity	Cognitive standard score	Documented ASD diagnosis	Discordance rules applied[a]	Change in Phase I status after consensus reveiw[b]
1	Never abstracted	Case	M	NH[c] black	68	No	Yes	NA
2	Never abstracted	Case	M	NH white	116	No	Yes	NA
3	Never abstracted	Case	F	NH black	25	No	No	NA
4	Never abstracted	Case	M	NH white	114	No	No	NA
5	Abstracted – non-case	Case	M	NH black	79	No	Yes	NA
6	Abstracted – non-case	Case	M	NH white	71	No	Yes	NA
7	Abstracted – non-case	Case	M	NH white	43	No	No	NA
8	Abstracted – non-case	Case	M	NH white	61	No	Yes	NA
9	Abstracted – non-case	Case	M	Hispanic	62	No	No	Yes
10	Abstracted – non-case	Case	F	NH white	63	Yes	Yes	Yes
11	Abstracted – non-case	Case	M	NH white	28	No	Yes	Yes
12	Abstracted – non-case	Case	M	NH white	111	No	No	Yes
13	Abstracted – case	Not case	M	NH white	91	Yes	No	NA
14	Abstracted – case	Not case	F	NH white	89	No	No	NA
15	Abstracted – case	Not case	M	Other	49	No	No	NA
16	Abstracted – case	Not case	F	NH black	100	Yes	No	NA
17	Abstracted – case	Not case	F	NH white	94	Yes	No	NA
18	Abstracted – case	Not case	F	NH black	63	No	No	No
19	Abstracted – case	Not case	M	NH white	127	Yes	No	NA

[a] Discordance rules defined in section labeled "Phase I and Phase II case definition"

[b] Phase I records-review status based on consensus decision after a second and independent review of surveillance records; *NA*, not applicable and consensus review not requested

[c] *NH*, non-Hispanic

the cases resolved by discordance rules fell into this misclassified group.

There were 12 false negatives and 7 false positives in the misclassified group (see Table 2). Specifically, 4 children defined as not abstracted/non-ASD cases in Phase I were defined as an ASD case in Phase II, 8 children defined as abstracted/non-ASD cases in Phase I were defined as ASD cases in Phase II, and 7 children defined as abstracted/cases in Phase I were defined as non-ASD cases in Phase II. Among the 19 children misclassified by Phase I records review, the following diagnoses were noted in surveillance records either in addition to or at the exclusion of an ASD: various medical conditions (e.g., mitochondrial disorder, $n = 10$), language disorders ($n = 8$), global developmental delays ($n = 7$), motor delays (including cerebral palsy, $n = 6$), attention deficit hyperactivity disorder ($n = 5$), learning disability ($n = 3$), anxiety disorders ($n = 3$), hearing loss ($n = 2$), seizure disorder ($n = 1$), and intellectual disability ($n = 1$). Other characteristics of the 19 children misclassified in Phase I of the study are outlined in Table 2.

Based on the ratio of ASD prevalence estimates comparing Phase II to Phase I, we estimated the mean adjustment to be an increase in prevalence of approximately 0.32

(CI$_{.95} = -0.04$, 0.68). The width of the confidence interval reflects the large sampling variability associated with the estimated surveillance parameters.

Secondary Analyses

Of the 39 children defined as Phase II clinical cases, 27 met absolute ADOS and ADI-R criteria, while 12 Phase II cases were included by discordant rules. When Phase II was based on the narrower case definition to meet both ADOS and ADI-R criteria absolutely, sensitivity increased 15% (from 60 to 75%); however, this was offset by a 14% reduction in the PPV (from 79 to 65%; see Table 1).

Table 3 details the 39 Phase II ("gold standard") ASD positives. Most of the 27 children who were also identified using Phase I surveillance methodology were those with a previous ASD diagnosis recorded by one or more community providers.

Discussion

Providing accurate prevalence estimates is a critical step toward understanding the impact and public health action

Table 3 Autism spectrum disorders (ASD) surveillance case classification and previously documented ASD diagnosis for all children determined to be ASD cases after Phase II clinical evaluation

		Previous ASD diagnosis documented in child's health or education records by community professionals		
		Yes	No	Totals
Phase I status	Case	23	4	27
MADDSP surveillance	Not case	0	12	12
ASD case classification	Total	23	16	39

MADDSP, Metropolitan Atlanta Developmental Disabilities Surveillance Program

needed to address ASD in the population. Results from this study demonstrate acceptable rates of the measures we used—specificity, PPV, and NPV—although the sensitivity rate was lower than anticipated. This is the first study to estimate the sensitivity of a population-based surveillance system; therefore, a direct comparison with other surveillance methods is not feasible. However, given our two-staged sampling design, where Phase I is treated as a screener for Phase II, we compared our results to those found in studies using clinical screens to detect ASD in school-aged children with special education needs. For example, Charman et al. (2007) found that the Social Communication Questionnaire (SCQ) had an estimated specificity of 0.78, PPV of 0.74, and NPV of 0.88 when administered in a population of children who required additional support in public school. Our Phase I surveillance screening estimates were comparable to the results of the SCQ screen, in that we found an estimated specificity of 0.96, PPV of 0.79, and NPV of 0.91. Yet, the sensitivity estimate for MADDSP was lower than that for the SCQ, reported by Charman et al. (2007) which found an estimated sensitivity of 0.86 (CI$_{.95}$ = 0.51, 0.94), compared with our estimated sensitivity of 0.60. However, the sensitivity in the former study was reduced to 0.77 when the sample was limited to children without intellectual disability. Further, the sensitivity of the SCQ in our sample was less than that derived from our a priori Phase II case definition (SCQ sensitivity = 0.51; CI$_{.95}$, = 0.38, 0.64) and both the SCQ and Phase II case definition were associated with wide CIs surrounding the sensitivity estimate. Thus, we suggest that our sensitivity estimate does not vary widely from clinical screens, and our point estimate is included in the range of confidence reported by other studies.

The wide CIs found in this study and other ASD screening studies suggest that the sensitivity of both clinical screens and records-review surveillance methods should be interpreted with caution. Further, the width of the CIs associated with the estimates implies that detection of children with the broad range of ASD is difficult despite the methodology employed. These conclusions may be particularly relevant for samples of children without intellectual disability and without a clear presentation of symptoms that warrant an ASD diagnosis. Another factor that can contribute to wide CIs is the use of small sample sizes to derive sensitivity estimates.

The importance and difficulty of identifying all cases of ASD, regardless of documented diagnosis, was demonstrated in one of our secondary analyses. We should not discount the fact that the Phase I records-review method did correctly identify some children without a documented ASD diagnosis (25%) who might be unaccounted for by other surveillance methods (e.g., review of service registries that rely on specific diagnostic codes for counting cases; Croen et al. 2002 and Hertz-Picciotto and Delwiche 2009). Conversely, our data also suggest that record-review surveillance methods might result in just as many false positives as true positives in classifying children without an ASD diagnosis as surveillance cases. Another limitation of the records-review method is that data are dependent on accessible sources, the accuracy and reliability of content, and the clarity of descriptors (e.g., phrases like "impairment in joint attention" could indicate a social initiation or response deficit). Given that some of the surveillance ASD cases were not previously identified in accessed records by a community professional, this may again highlight challenges of identifying all children with ASD using population-based designs, even when behavioral symptoms are present. However, searching for children with the broad array of ASD symptoms is a critical aspect of any ASD surveillance system and is a potential strength of records-based approaches, compared with other approaches that do not make provisions for children with undiagnosed ASD. This point is highlighted by surveillance methods that rely on only clinical diagnosis and potentially underestimate the number of children with ASD in the population (Barbaresi et al. 2009; Centers for Disease Control and Prevention 2009).

Several children misclassified as not having an ASD during Phase I records-review would have been classified as an ASD case by MADDSP, given a coding rule change that was implemented in the surveillance year after start of

this study. The change concerns the transfer of a surveillance case to non-case status, based on consensus decision and adjustments in prevalence based on surveillance files not located at partner sources (and thus not available for data collection for clinician review). Even when these factors were taken into account, the adjusted sensitivity, specificity, PPV, and NPV estimates (data not shown) were similar to the original estimates; yet, several more misclassified children were incorrectly transferred from case to non-case status or were missing health or education files.

Several limitations in this study deserve consideration. First, the Phase II case definition used to estimate sensitivity, specificity, PPV, and NPV is also associated with imperfect measures of the aforementioned parameters (Lord et al. 1989, 1994; Risi et al. 2006), and we used rules to resolve discordance between ASD diagnostic measures that have not been studied or replicated in published research. Thus, the decision to include children who did not meet full diagnostic criteria on one of these instruments may have negatively impacted study results. The increase in sensitivity when the Phase II clinical case definition was refined (and did not include resolved discordant cases) speaks again to the uncertainty of the estimate and the difficulty detecting children with an ASD who do not have a clear presentation of symptoms. It also highlights the variability associated with applying different interpretations of diagnostic test results. Under the first set of criteria we end up with a moderate sensitivity but fairly high PPV, while with the restricted set of assumptions, the sensitivity increases but the PPV decreases.

Another limitation might be our failure to recruit the target sample for the not abstracted/non-case group during Phase II clinical evaluations. The reduced sample size could be indicative of a selection bias, in that families who suspect their child may have an ASD are more likely to participate in the clinical evaluation. However, simulation analyses suggested that less-than-optimal recruitment, while increasing the sampling variability associated with the estimates, was not likely to result in substantial bias in the estimates even when we assumed that the true prevalence of ASD cases was two times lower among children whose cases were never abstracted or whose cases were abstracted/non-ASD cases (during Phase I records-review).

Finally, the average cognitive standard score in our sample was 89, and 72% of participants did not have intellectual disability. Consequently, the cognitive characteristics of our sample may not reflect all children with ASD in the population and could have negatively influenced the sensitivity estimate (Centers for Disease Control and Prevention 2009). Children without cognitive impairment might not receive special education services, and are therefore more difficult to ascertain through these records-review surveillance methods.

Our study offers several important strengths. Ours is the first analysis of a records-based surveillance system, and we found rates of specificity, PPV, and NPV comparable to clinical screens in similar samples of children. Our estimate of sensitivity was reduced, although the uncertainty of this estimate is striking and strongly influenced by the clinical case definition employed and presence of a documented ASD diagnosis. Importantly, one of the primary aims of a surveillance system is to identify overall prevalence of a condition. In our study, the total number of children with missed ASD cases is only 5 children, if we consider the net difference in false positives and false negatives. Given this small number of net missed-ASD cases, coupled with the uncertainty surrounding our sensitivity estimate, prevalence adjustments based on these data may not be warranted; however, if the case definition is applied to larger surveillance populations in future years, the proportion of missed children may be much greater and should be monitored.

Since ASD is a group of heterogeneous and behaviorally-defined disorders, any attempt to identify children with the broad array of ASD is challenging even with established gold standard diagnostic tools. Resolution of discordant ADOS and ADI-R diagnostic algorithms is a critical aspect of case determination. As evidenced by this study, the clinical confirmation of surveillance cases without a documented diagnosis supports the strengths of MADDSP methods, despite less than ideal sensitivity. Thus, one might consider that records-based approaches may provide more accurate prevalence estimates than surveys based on parental reports or examination of service registries, because records-based approaches do not rely on a documented ASD diagnosis to calculate prevalence. Further, records-based approaches are more feasible than population-based clinical screening and evaluation. Even so, continual refinement of records-based surveillance systems is needed to obtain the most accurate estimates of ASD prevalence to best inform public health needs. Consequently, our recommendations for records-based surveillance programs are to continue to refine coding rules and provide more resources to search for missing files. Coding rules that allow a child to be transferred from surveillance case to non-case status based on consensus decision should be given particular attention in future surveillance years.

References

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

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). *Lancet, 368*, 210–215.

Barbaresi, W. J., Colligan, R. C., Weaver, A. L., & Katusic, S. K. (2009). The incidence of clinically diagnosed versus research identified autism in Olmstead County, Minnesota, 1976–1997: Results from a retrospective, population-based study. *Journal of Autism and Developmental Disorders, 39*, 464–470.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). The autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry, 175*, 444–451.

Boyle, C. A., Yeargin-Allsopp, M., Doernberg, N., Holmgreen, P., Murphy, C. C., & Schendel, D. (1996). Prevalence of selected developmental disabilities in children aged 3–10 years: The Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1991. *MMWR Surveillance Summaries, 45*((No. SS-2)), 1–14.

Centers for Disease Control and Prevention. (2009, December 18). Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, United States, 2006. *Surveillance Summaries MMWR, 58*(No. SS-10).

Centers for Disease Control and Prevention, ADDM Network Surveillance Year 2002 Principal Investigators. (2007, February 9). Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, 14 sites, United States, 2002. *MMWR Surveillance Summaries, 56*(SS01, 2007), 2–28.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., et al. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry, 191*, 554–559.

Croen, L., Grether, J., Hoogstrate, J., & Selvin, S. (2002). The changing prevalence of autism in California. *Journal of Autism and Developmental Disorders, 32*, 207–215.

Elliot, C. D. (1990). *The differential ability scales*. San Antonio, TX: Harcourt Assessment.

Filipek, P., Accardo, P., Ashwal, S., Baranek, G. T., Cook, E. H., Dawson, G., et al. (2000). Practice parameter: Screening and diagnosis of autism. *Neurology, 55*, 468–477.

Fombonne, E. (1996). Is the prevalence of autism increasing? *Journal of Autism and Developmental Disorders, 26*, 673–676.

Gillberg, C., Steffenburg, S., & Schaumann, H. (1991). Is autism more common now than ten years ago? *British Journal of Psychiatry, 158*, 403–409.

Hertz-Picciotto, I., & Delwiche, L. (2009). The rise in autism and role of age at diagnosis. *Epidemiology, 20*, 84–90.

Johnson, C. P., Myers, S. M., & American Academy of Pediatrics Council on Children with Disabilities. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*, 1183–1215.

Kadesjo, B., Gillberg, C., & Hagberg, B. (1999). Brief report: Autism and Asperger syndrome in seven-year-old children: A total population study. *Journal of Autism and Developmental Disorders, 29*, 327–331.

Kawamura, Y., Takahashi, O., & Ishii, T. (2008). Reevaluating the incidence of pervasive developmental disorders: Impact of elevated rates of detection through implementation of an integrated system of screening in Toyota, Japan. *Psychiatry and Clinical Neurosciences, 62*, 152–159.

Kogan, M. D., Blumberg, S. J., Scheive, L., Boyle, C., Perrin, J., Ghandour, R., et al. (2009). Prevalence of parent reported diagnosis of autism spectrum disorder among children in the U.S., 2007. *Pediatrics, 124*, 1395–1403.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). *Autism diagnostic observation schedule*. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., et al. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behaviors. *Journal of Autism and Developmental Disorders, 19*, 185–212.

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*, 659–685.

McNamee, R. (2002). Optimal designs of two-stage studies for estimation of sensitivity, specificity and positive predictive value. *Statistics in Medicine, 21*, 3609–3625.

Posserud, M., Lundervold, A., & Gillberg, C. (2006). Autistic features in a total population of 7–9-year-old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *Journal of Child Psychology and Psychiatry, 47*, 167–175.

Rice, C., Baio, J., Van Naarden Braun, K., Doernberg, N., Meaney, F. J., Kirby, R. S., et al. (2007). A public health collaboration for the surveillance of autism spectrum disorders. *Paediatric and Perinatal Epidemiology, 21*, 179–190.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*, 1094–1103.

Rutter, M. (2005). Incidence of autism spectrum disorders: Changes over time and their meaning. *Acta Paediatrica, 94*, 2–15.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R: The autism diagnostic interview-revised*. Los Angeles, CA: Western Psychological Services.

Schieve, L., Rice, C., & Boyle, C. (2006). Mental health in the United States: Parental report of diagnosed autism in children aged 4–17 years, United States, 2003–2004. *MMWR, 55*, 481–486.

Sparrow, S., Balla, D., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Services.

Van Naarden Braun, K., Pettygrove, S., Daniels, J., Miller, L., Nicholas, J., Baio, J., et al. (2007, February 9). Evaluation of a methodology for a collaborative multi-source surveillance network for autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2002. *MMWR Surveillance Summaries, 56* (No. SS-01), 29–40.

Zhou, X. (1998). Correcting for verification bias in studies of a diagnostic test's accuracy. *Statistical Methods in Medical Research, 7*, 337–353.