ORIGINAL PAPER

What Drives Detection and Diagnosis of Autism Spectrum Disorder? Looking Under the Hood of a Multi-stage Screening Process in Early Intervention

R. Christopher Sheldrick[2,6] · Elizabeth Frenette[1] · Juan Diego Vera[1] · Thomas I. Mackie[3,4] ·
Frances Martinez-Pedraza[5] · Noah Hoch[1] · Abbey Eisenhower[1] · Angel Fettig[1] · Alice S. Carter[1]

Abstract

U.S. guidelines for detecting autism emphasize screening and also incorporate clinical judgment. However, most research focuses on the former. Among 1,654 children participating in a multi-stage screening protocol for autism, we used mixed methods to evaluate: (1) the effectiveness of a clinical decision rule that encouraged further assessment based not only on positive screening results, but also on parent or provider concern, and (2) the influence of shared decision-making on screening administration. Referrals based on concern alone were cost-effective in the current study, and reported concerns were stronger predictors than positive screens of time-to-complete referrals. Qualitative analyses suggest a dynamic relationship between parents' concerns, providers' concerns, and screening results that is central to facilitating shared decision-making and influencing diagnostic assessment.

Keywords Autism spectrum disorder · Screening · Costs · Decision-making · Process assessment

Abbreviations

ASD	Autism spectrum disorders
EI	Early intervention
IOM/NAM	Institute of Medicine/National Academy of Medicine
BITSEA	Brief infant and toddler social emotional assessment
POSI	Parent observation of social interaction
STAT	Screening tool for autism in toddlers

✉ R. Christopher Sheldrick
rshldrck@bu.edu

1 Department of Psychology, University of Massachusetts Boston, Boston, MA, USA

2 Department of Health Law, Policy and Management, School of Public Health, Boston University, One Silber Way, Boston, MA 02215, USA

3 Rutgers School of Public Health, Piscataway, NJ, USA

4 Institute for Health, Health Care Policy and Aging Research, Rutgers University, New Brunswick, NJ, USA

5 Department of Psychology, Florida International University, Miami, FL, USA

6 Boston University, 715 Albany Street, Boston, MA 02118, USA

Introduction

Improving identification and treatment of Autism Spectrum Disorder (ASD) requires not only accurate screening, but also effective systems to ensure that children at risk are appropriately assessed, diagnosed and treated. At a systems level, pediatricians in the U.S. who are concerned about ASD and/or developmental delays often refer to Early Intervention (EI) for further services, which typically address developmental delays. Although risk for ASD is high among children with specific or global developmental delays (Wiggins et al. 2015), differential diagnoses of ASD are often missed or delayed once referred to EI. Given that recommended and reimbursable pediatric visits for young children occur periodically with lags ranging from 3 to 12 months (King et al. 2010), pediatric providers are not always able to fully monitor the emergence of ASD diagnoses that can occur during this phase of development (Ozonoff et al. 2015). Improving identification of ASD in EI settings can therefore enhance connections between screening in primary care and downstream diagnoses and ASD-specific interventions.

At the process level, how best to ensure that individual children who are referred to EI are appropriately screened and evaluated is an open question. While evidence-based

screening instruments are widely recommended, they are seldom implemented in EI settings. In addition, connecting children to appropriate services depends on at least two elements: (1) an effective decision rule to determine which children demonstrate sufficient risk to warrant follow-up (i.e., risk stratification), and (2) family engagement with recommended screening, assessment, and intervention services. These two elements imply that the presentation and results of the screening process should be convincing enough to influence the judgment and decision-making of both the providers, who make referrals, and the parents, who choose whether or not to engage with recommended services.

In this paper, we report the results of a mixed-methods evaluation designed to "look under the hood" of a multi-stage targeted screening process in EI. We have two specific aims: (1) to evaluate a clinical decision rule that actively encourages referral to the next stage of assessment not only if screening results are positive, but also if concern is reported by parents or EI providers, and (2) to characterize the process of shared decision-making between parents and EI providers in which screening occurs.

Clinical Decision Rules

Professional guidelines for screening, which focus on primary care settings, conventionally recommend the clinical decision rule that all children who screen positive be referred for evaluation or treatment. Concordantly, there is an extensive body of research that focuses on this element of screening, attempting to assess the accuracy of specific screening instruments (Zwaigenbaum et al. 2015). However, professional guidelines also make recommendations about when and the degree to which clinicians should exercise their professional judgement. However, these recommendations are less standardized, and screening research often overlooks the details of clinical judgment, such as how professionals integrate evidence gathered through the process of screening and observation to arrive at an informed opinion regarding their degree of clinical concern. For example, the American Academy of Pediatrics (AAP) policy for identifying children with developmental disorders stipulates that an evidence-based screening tool should be administered at specific visits or be triggered by concerns noted in the course of standard pediatric care; therefore, referrals for diagnostic assessment are mostly recommended on the basis of positive screening results (Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee & Medical Home Initiatives for Children With Special Needs Project Advisory Committee 2006). In contrast, the AAP policy for identifying children with ASD recommends referrals based either on positive screening results or on the presence of at least two risk factors, such as having a sibling with ASD and there being concern

from either a provider and/or caregiver (Johnson and Myers 2007). While similar in many ways, the differences between the two policy statements have significant implications for whether and how clinicians should incorporate screening results into their clinical decision-making.

Recommendations regarding clinical judgment have a long and controversial history. There is a longstanding tradition in psychological research of basing clinical decisions solely on the results of quantitative instruments, with some investigators even suggesting no room at all for clinical judgment (Swets et al. 2000). Several meta-analyses have concluded that quantitative prediction tools (referred to in these studies as actuarial methods) are superior to clinical judgment in reducing total error rates (i.e., total number of false positive plus false negative errors) (Ægisdóttir et al. 2006; Grove et al. 2000; Meehl 1959). Subsequent reviews have disseminated these findings and struggled to explain why clinicians are resistant to their conclusions (Dawes et al. 1989; Grove and Meehl 1996). As one example, an article published in Science in 1989 states that "research comparing these two approaches shows the actuarial method to be superior…failure to accept a large and consistent body of scientific evidence over unvalidated personal observation may be described as a normal human failing or, in the case of professionals who identify themselves as scientific, plainly irrational" (Dawes et al. 1989). From this actuarial perspective, the predictive accuracy of tests and other quantitative methods is of paramount concern. It is therefore not surprising that the Institute of Medicine/National Academy of Medicine (IOM/NAM) has observed that "the majority of scientific evidence about any diagnostic test typically is focused on test accuracy," rather than on its optimal use (Balogh et al. 2015; Brożek et al. 2009; Trikalinos et al. 2009). On this subject, psychology stands apart from other fields, including medicine, which have long acknowledged the importance of clinical judgment (Balogh et al. 2015).

The role of clinical judgment in the detection of ASD is particularly complex. In addition to inclusion of concern into AAP recommendations, numerous studies document that, in practice, primary care providers often depart from professional guidelines recommending referral of all children who screen positive (Robins et al. 2001, 2013), both for ASD and for other developmental-behavioral conditions (Sheldrick et al. 2016). Recent research demonstrates that such decisions are not inherently irrational given the modest negative and positive predictive values of most screening instruments—especially for scores near the clinical decision threshold, or "the cut score" (Sheldrick and Garfinkel 2017; Sheldrick et al. 2015). In light of this understanding of decision thresholds, the above-cited meta-analyses that compare clinical judgment to quantitative prediction tools have a clear limitation. The vast majority of the papers included in these meta-analyses

report total error rates but fail to distinguish between false positive and false negative errors as well as the differential real-world impacts of those errors. A decision threshold that is optimal with respect to total error rate is not necessarily a decision threshold that is optimal with respect to risks and benefits. A risk/benefit optimization often yields a higher number of errors than a total error rate optimization. For example, consider a screening algorithm for ASD that is optimized with respect to the total error rate. Such an algorithm would be likely to yield equal numbers of false positives (i.e., children without ASD referred for evaluation) and false negatives (i.e., children with ASD who are not referred). Notably, the primary author of the original meta-analysis cited above assumes that such a screening algorithm is optimal without considering whether it is preferable to other options (Meehl and Rosen 1955), such as one with a lower threshold that yields more errors overall but misses fewer children with ASD (i.e., fewer false negatives but more false positives). A large body of research on decision analysis suggests that the optimal ratio of false positive to false negative results depends on the relative costs associated with each (Pauker and Kassirer 1980). By focusing on minimization of total error rates, the meta-analyses cited above neglect to question if false positive and false negative errors have equal impact. A more recent meta-analysis comparing clinical judgment to quantitative instruments in medicine included papers that distinguished between false positive and false negative results and found no clear differences in overall accuracy (Sanders et al. 2015). Thus, conclusions regarding the established superiority of actuarial methods to clinical judgment are open to question.

Despite widespread use of clinical judgment in practice and its inclusion in professional recommendations, there is little published evidence regarding the use of clinical judgment in decision-making. Indeed, many studies treat ASD screening and clinical judgment as mutually exclusive processes. For example, a recent policy statement from the United States Preventive Services Task Force (USPSTF) concluded that systematic, ASD-specific screening was more effective in identifying children with ASD early than relying on "developmental surveillance, general developmental screening, and/or parental concerns" (McPheeters et al. 2015). Thus, while the USPSTF was able to identify seventeen studies that evaluated the accuracy of ASD-specific screening instruments in unselected populations, their review identified no studies that specifically evaluate the integration of clinical concern with screening results when making referral decisions (McPheeters et al. 2015). Given the AAP's policy on integrating clinical judgment and evidence-based screening instruments when making referral decisions for ASD (Johnson and Myers 2007), it is important to conduct research that focuses on how clinicians use screening instruments in practice and whether clinical judgment helps or hinders the process of identifying children with ASD.

In this study, we evaluate a decision rule for combining clinical judgment with evidence-based screening implemented in EI settings. Our decision rule specified that children could be referred for stage 2 screening based on either a positive screen and/or report of concern from the parent or EI provider. Note that because concerns were assessed at the end of the screening forms, they may have either arisen either before or because of screening. Because children in our study population have qualified for EI and are therefore already at risk for ASD, we required only one report of additional concern in our first stage of screening rather than the two required by AAP guidelines for the detection of ASD in primary care (Johnson and Myers 2007).

Shared Decision-Making

In its recommendations for reducing diagnostic errors in medicine, the IOM/NAM emphasizes not only diagnostic test accuracy, but also shared decision-making between clinicians and patients to determine whether assessed risk is sufficient to justify further action (Balogh et al. 2015). For example, even if the raw score of an ASD screening instrument offers a valid indicator of risk (i.e., with higher scores reflecting increased predictive value across the full range of scores), parents' opinions may still differ as to whether that risk is sufficient to justify further evaluation. In evaluating the effectiveness of a screening protocol, a focus on shared decision-making is essential because while providers typically decide whether to refer for services, it is parents who decide whether to engage in those services (including the screening itself). Both provider and parent engagement are essential to ensure that children with ASD receive appropriate care.

The importance of shared decision-making has at least two implications for the present study. First, successful referrals that result in further screening or evaluation should be interpreted as reflecting not only fidelity to the protocol or a clinical decision by the provider but also as a choice by the parent. Second, the process by which EI providers use and interpret screening instruments is likely to be informed by interactions with parents. In the research literature, screening instruments are typically conceptualized simply as indicators of risk, with positive scores indicating a need for referral. While EI providers may use screening instruments to gather information to inform their own decisions, they must also consider parents' responses to screening. Parents' reactions can be both subtle and complex, including whether to engage in screening and how much to disclose during the screening process. Thus, how EI providers use screening instruments is likely to be informed by their perceptions of

parents' concerns and their appraisals of the family's readiness to engage in the screening process, which involves shared decision-making.

To better understand the relative contributions of screening results and concerns under our screening protocol, we analyze the value of positive screening results, EI provider concern, and EI providers' perceptions of parent concerns at baseline for predicting successful completion of subsequent referrals for assessment. To explore the interconnections between screening and shared decision-making in greater depth, we complement these analyses with qualitative evidence derived from interviews and participant observations with parents and EI providers.

Through the evaluation of our decision rule and the exploration of shared decision-making, results offer a "look under the hood" of a multi-stage targeted screening process designed to improve linkage to services in a population of children who were initially identified as at risk and referred to EI.

Methods

This study is part of a community-based research project that utilized a Type II effectiveness-implementation hybrid approach to reduce health disparities in access to ASD services (Curran et al. 2012). To address the current study aims, we employed a convergent, mixed methods design with concurrent quantitative and qualitative components (Creswell et al. 2011). Aim 1 was to evaluate a clinical decision rule that actively encourages referral to the next stage of assessment not only if screening results are positive, but also if concern is reported by parents or EI providers. To address this aim, we used quantitative analyses to evaluate the effectiveness of our decision rule by comparing participation rates, rates of confirmatory results (i.e., positive screens and diagnostic evaluations), and program cost among children referred for further evaluation based on three different pathways: (Path A) a positive screening test and report of concern, (Path B) report of concern alone [no positive screening test], (Path C) a positive screening test alone [no report of concern].

Aim 2 was to characterize the process of shared decision-making between parents and EI providers in which screening occurs. To do so, we used mixed methods to holistically analyze the influence of baseline screening results and presence of concerns on completion of the stage 2 screening and/or diagnostic evaluation. To address this aim in quantitative terms, we conducted a survival analysis to determine the predictive value of positive screening scores, EI provider concern, and provider perception of parent concern at baseline for completion of stage 2 screening. To address this aim in qualitative terms, we conducted qualitative analyses

of observations and semi-structured interviews with parents and EI providers to better understand how the process of shared decision-making varies depending on the presence or absence of concern. Where appropriate, procedures are described separately for each component.

Participants

Quantitative Component

Participants were parents of children who were eligible for ASD screening between November 1 2014 and February 14 2018 at three EI agencies that primarily serve low-resource families in an urban, New England area of the USA, as well as a sub-sample of EI providers from these agencies. For children 0–3 years of age and their families, EI agencies provide IDEA Part C family-centered services to address problems in five domains: speech and language, social emotional skills, motor skills, adaptive, and cognitive development. At each site, our research team offered training and support to EI providers to implement a multi-stage screening process designed to detect and diagnose ASD. The screening process was incorporated by each site as part of their routine clinical practice. Participating families who screened positive at both screening stages 1 and 2 were offered a free diagnostic evaluation conducted by clinical psychologists on the research team. Families were eligible for screening if their children (a) were enrolled in EI, (b) had no previous diagnosis of ASD, (c) had no medical condition that would limit the ability to diagnose ASD, (d) were between the ages of 14 and 33 months, and (e) a parent or guardian understood Spanish or English sufficiently to complete questionnaire screeners.

Qualitative Component

A sub-sample of EI providers from one of the three partner agencies participated in qualitative interviews. Comparable to sociodemographic characteristics of the larger EI provider populations at these urban EI agencies, all twenty EI providers who completed qualitative interviews identified as female, seventeen self-identified racially/ethnically as White, two as Hispanic, and one as Asian-American. The median age of respondents was 28 years (range 24–42). All respondents completed at least 4 years of college, and fourteen held Master of Arts or Education degrees in fields related to their current work in EI, such as occupational therapy, speech and language therapy, and child development. Signaling the diverse needs of the population they serve, eight respondents reported experience in delivering services in Spanish and two respondents in American Sign Language.

In addition, we conducted approximately 40 hours of observations of the multi-stage screening protocol and 63

interviews with 22 parents in a longitudinal qualitative sub-study. Nineteen of the parents identified as female, and within that group, nine identified as African-American, five identified as White, one as North-African, one as Asian, and three as racially Black and ethnically Hispanic. Three of the parents identified as male, and of that group, two identified as White and one identified as Hispanic.

Multi-stage Screening Protocol

The screening process consisted of three stages. Stage 1 questionnaires and stage 2 observational screenings were administered by EI providers as part of their routine clinical practice, whereas research staff conducted stage 3 diagnostic assessments at a university clinic.

Stage 1 Screening

Parents completed two screening instruments: the Brief Infant Toddler Social Emotional Assessment (BITSEA) and the Parent's Observations of Social Interactions; (POSI). The BITSEA is a global screener of social emotional development (Briggs-Gowan and Carter 2006). The BITSEA includes an ASD-specific risk index that sums responses regarding 10 problem behaviors and subtracts responses regarding 9 competencies relevant to ASD; the resulting score displayed high accuracy in detecting ASD diagnoses in a previous study (Giserman Kiss et al. 2017). The POSI is a second ASD-specific screener that has demonstrated strong sensitivity and adequate specificity in two previous studies (Salisbury et al. 2018; Smith et al. 2013). Two screening instruments were included to maximize sensitivity and minimize false negative results. EI Providers introduced the screening packet to parents, scored the measures, and then recorded whether or not the following two statements were true: "EI team concerned about ASD" and "Parent concerned about ASD." A positive response on either question was coded as indicative of concern for purposes of referral. Families were referred to stage 2 screening if: (1) children scored positive on either of the two BITSEA ASD indices (i.e., ASD Problem or ASD Competence) or the POSI, or (2) EI Providers reported either their own or parental concern about ASD. Following evidence from previous studies that provider concern could generate additional referrals (Robins et al. 2013), the explicit goal of this decision rule was to maximize the sensitivity of the initial screening.

Stage 2 Screening

With parents and their primary EI providers present, specially-trained EI providers administered a 20-min, play-based observational assessment called the Screening Tool for ASD for Toddlers and Young children (STAT) (Stone et al. 2000, 2008).

Families were referred for a diagnostic assessment if: (1) children scored positive on the STAT, or (2) EI Providers reported their own or parental concern about ASD.

Stage 3 Diagnostic Assessments

With parents present and often accompanied by an EI provider, diagnostic assessments were administered in a university clinic by post-baccalaureate research assistants or doctoral students in Clinical Psychology who were trained to research reliability on the ADOS-2 and observed by licensed psychologists with expertise in ASD in early childhood. Assessments included the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) (Lord et al. 2000), the Mullen Scales of Early Learning (Mullen 1995), and the parent/caregiver interview form of the Vineland Adaptive Behavior Scales, Third Edition (Sparrow et al. 2016). Final DSM-5 ASD diagnoses were assigned based on the clinician's evaluation after observing the visit and reviewing all of the diagnostic assessment results.

EI Provider Training

The research team provided ongoing training to EI providers in the use of the two evidence-based stage 1 instruments that screen for signs and symptoms of ASD. Training focused on both technical aspects of administration as well as how to open and maintain often difficult conversations with parents about the presence of risk for ASD. To accommodate turnover among EI staff, both live trainings and training videos were provided. At each site, a small team of EI providers was trained to administer the stage 2 screening instrument, the STAT. Typically, both the primary EI provider and the specially-trained STAT EI provider discussed the results with the parents. Following a positive stage 2 screening (i.e., positive STAT or ongoing parent or provider concern), the EI provider could use a web-based application to schedule a diagnostic evaluation. The EI providers were also given guidelines for discussing results and concerns with families to use throughout the screening process. In addition, the primary EI provider routinely accompanied families to the diagnostic assessment that was conducted by the research team. These procedures were designed to support families, improve EI providers' sensitivity to the early signs and symptoms of ASD, and engage both parties in the process of shared decision-making across the multi-stage screening protocol.

Study Procedures

Quantitative Component

Quantitative data were collected to compare the success of referrals for stage 2 screening based on three different

pathways defined by stage 1 screening results. Specific outcomes included participation rates at each stage (i.e., the number of families who completed 2nd stage screening and diagnostic assessments among those referred) and rates of confirmatory results at each stage (i.e., the number of families who received positive, "of concern" screening scores or ASD diagnoses among those who completed assessments).

To assess the utility of including a decision rule that encourages referrals based on concern alone, the marginal cost to the program of assessing children in each pathway in regard to provider time expenses was assessed. To do so, research staff estimated the time to complete stage 2 screening and diagnostic evaluations by directly observing 14 stage 2 visits. Hourly cost was estimated based on reimbursement rates offered by the state's Medicaid program. For each referral pathway, the number of stage 2 screens conducted and the number of diagnostic evaluations conducted were counted. Hours per assessment and number of assessments were multiplied to estimate the total staff hours required to evaluate children in each pathway, and the result was then multiplied by cost/hour to estimate cost. These estimates reflect the marginal cost to the program of assessing children in each pathway. To estimate benefit to the program, the number of ASD diagnoses yielded by each pathway were counted. Dividing the number of ASD diagnoses by the marginal cost offers an indicator of the marginal cost-effectiveness of including each pathway in our decision rule.

We hypothesized that Path A (positive screen + concern) would be most efficient, yielding the lowest cost per ASD diagnosis. Because Path C (positive screen with no concern) is included in most professional screening recommendations, comparison with Path B (negative screen + concern), offers a useful evaluation of the relative cost/benefit of including clinical judgement in our decision rule.

Qualitative Component

Brief surveys and an in-person semi-structured qualitative interview protocols were developed by trained qualitative researchers for both EI providers and parents; interview guide domains and measures were informed by those published in the peer-reviewed literature and an inter-disciplinary research team. First, the EI provider survey and interview guide consisted of 40 categorical and open-ended questions that characterized respondents' (1) sociodemographic characteristics (Stahmer 2007; Stahmer et al. 2005), (2) experiences providing EI services(Stahmer 2007; Stahmer et al. 2005), (3) experiences administering ASD screening tools specifically, and perceptions of ASD (Pizur-Barnekow et al. 2013; Pizur-Barnekow and Schefkind 2014). Reported in this paper, EI providers were asked to describe the multi-stage screening process, and follow-up questions included specific probes about strategies they used when

they and/or the parent lacked concern of ASD and rationale for these actions. Second, parents were asked about their perspectives on the ASD screening process also drawing on domains and measures in extant literature and interdisciplinary team review, including (1) sociodemographic characteristics (Guinchat et al. 2012; Zuckerman et al. 2014a), (2) perceptions of developmental and ASD concerns (Guinchat et al. 2012; Zuckerman et al. 2014a, b), (3) perceptions of screening tools and results (Calzada et al. 2012; Zuckerman et al. 2014a), and (4) multi-stage screening protocol. Reported in this paper are parental descriptions of the multi-stage screening process and detailed data about their perspectives on shared decision-making, specifically how both the EI provider concerns' and their own influenced whether their toddler advanced (or did not advance) to subsequent stages of the multi-stage protocol.

Following the brief paper-and-pencil survey to collect sociodemographic factors, we conducted the semi-structured qualitative interviews with parents and EI providers. Interview protocols were developed by trained qualitative researchers with ongoing review from an interdisciplinary research team. Participants received monetary compensation for their travel and time. Lasting approximately 80 min, each interview was audiotaped and transcribed verbatim.

For providers and parents, our analyses achieved thematic saturation on core domains described above (i.e., no new information after approximately 15 parent and 15 provider interviews) indicating a sample of adequate size for qualitative analyses presented (Fusch and Ness 2015). All study procedures were approved by the institutional review board and informed consent was documented in writing.

Quantitative Analyses

Quantitative analyses were conducted using Stata version 15. Descriptive statistics were calculated for demographic and baseline variables for the total sample and by referral pathway, and regression analyses were conducted to test for differences.

Aim 1

To compare the utility of different referral pathways that result from our decision rule, participation and diagnostic rates for each screening to diagnostic evaluation pathway were calculated directly from project administrative data and compared using logistic regression followed by Wald tests. In addition, survival analyses using Cox regression models were conducted to compare time-to-completion from initial screening to the stage 3 diagnostic assessment among the three pathways. To account for partially missing data at stage 1 (i.e., children with screening results for whom EI providers did not record presence or absence of concern), analyses

were conducted both on raw data using case wise deletion and across 20 multiple imputations under the assumption that data were missing at random. Results of each analysis were compared for consistency. To assess whether results generalize across EI agencies, we tested whether effects were moderated by site. Because the analyses were designed to inform programmatic decisions regarding the inclusion of a specific decision rule, we did not adjust for factors like age, gender, and prior symptom scores. Instead, we offer descriptive data for children identified through each pathway that includes age, initial screening scores, and final ADOS scores. Likewise, our interest was in the utility of the overall screening process for identifying and diagnosing ASD. Because a missed case can result either from a false negative screening result or from a true case who drops out of a screening process, we did not adjust statistically for dropout.

Aim 2

To analyze the influence of baseline screening results and presence of concerns on completion of later screens, we first calculated descriptive statistics to reflect associations between screening scores, provider concern, perception of parent concern, and completion of stage 2 screening. In addition to histograms, we also conducted logistic regressions to calculate the expected probability of concern and of completion of stage 2 screening based on the BITSEA total ASD scale, a sum of the two ASD scales included in the protocol's decision rule that was published after the study began (Gisserman-Kiss et al. 2017). To further evaluate variables included in our decision rule, we conducted survival analysis of time-to-complete stage 2 screening using Cox regression. Independent variables included stage 1 positive screening results, EI provider concern, and EI provider perception of parent concerns.

Qualitative Analyses

Qualitative data from semi-structured interviews and participant observations were analyzed using a modified grounded theory approach known as "Coding Consensus, Co-occurrence, and Comparison," in which analyses are derived from the data and then illustrated by characteristic examples (Willms et al. 1990). To develop separate structured codebooks for parent and EI provider transcripts, transcripts of three interviews were independently coded by an interdisciplinary team of investigators. Investigators coded excerpts of the transcripts, ranging from a phrase to several paragraphs, based on *a priori* or emergent themes. Disagreements were discussed and resolved, strengthening the definition of the respective codes. Based on the drafted codebooks, three investigators separately reviewed two additional transcripts to determine level of agreement in the codes applied. A

final set of code definitions were then discussed, resolved, and recorded. All transcripts for both providers and parents were then reviewed, coded, and subsequently compared and reconciled by at least two of the investigators. Throughout this process, we used intensive group discussion as our goal was consensus rather than employment of quantitative measures of inter-rater agreement (Sturges and Klingner 2005). Coded data were entered into DeDoose, a mixed methods software program, and a series of categories for these data generated with links between the categories. Field observations were inductively coded emerging data on variations in administration based upon the extent of parental and provider concerns. For this paper, we analyzed codes capturing the extent of parental or provider concerns for ASD and considered these codes across the initiation, implementation of the multi-stage screening protocol, and subsequent next steps. As described above, the qualitative components of this study employed multiple methods (interview and participant observation) and respondent samples (EI provider and parents) allowing for methodological and respondent triangulation and thus improving the ability to achieve crystallization and completeness in our understanding of the process for shared decision-making (Tobin and Begley 2004).

Results

Aim 1: Quantitative Results

Demographic characteristics of the 1,654 children who completed stage 1 screening are described in Table 1. The sample was diverse with respect to race/ethnicity (72% who reported race/ethnicity identified as non-white and/or Hispanic or Latino), language (33% of families reported a primary language other than English), and insurance status (60% reported receiving public health insurance). Average child age at stage 1 screening was 23.8 months (SD = 5.4), and average age of ASD diagnosis was 27.4 months (SD = 4.9). Compared to children in Path A, children in Paths B and C had lower scores on the BITSEA ASD problems scale and higher scores on the BITSEA ASD Competence scale (i.e., fewer ASD-relevant problems and more competencies). Thus, concern was more likely to be reported for children with high screening scores than for children whose results were closer to the recommended thresholds, or cut scores. In addition, children in Path B were somewhat older (mean difference = 1.7 months), while children in Path C were more likely to be female. Table 1 also depicts results of diagnostic evaluations for children who completed them in each pathway. Compared to children in Path A, children in Paths B and C both scored higher on the Mullen Early Learning Composite, while children in Path B also scored higher on the Mullen Expressive Language scale and the

Table 1 Participant demographics (% or standard deviation)

	Total 1654 (100%)	Pathway A 311 (19%)	Pathway B 121 (7%)	Pathway C 183 (11%)	Not referred 1039 (63%)	p value	
Stage 1 ASD screen		+	−	+	−	B versus A	C versus A
Clinical concern		+	+	−	−		
Gender						−	**
Female	588 (36%)	57 (18%)	27 (22%)	68 (37%)	436 (42%)		
Male	1048 (63%)	250 (80%)	93 (77%)	114 (62%)	591 (57%)		
Not reported	18 (1%)	4 (1%)	1 (1%)	1 (1%)	12 (1%)		
Race/ethnicity						−	−
White (non-Hispanic)	297 (18%)	40 (13%)	18 (15%)	28 (15%)	211 (20%)		
Black or African American	307 (19%)	69 (22%)	32 (26%)	26 (14%)	180 (17%)		
Black hispanic	96 (6%)	31 (10%)	9 (7%)	10 (5%)	46 (4%)		
Hispanic/latino	361 (22%)	92 (30%)	34 (28%)	40 (22%)	195 (19%)		
Asian	66 (4%)	23 (7%)	4 (3%)	4 (2%)	35 (3%)		
Multiracial/multiethnic	81 (5%)	14 (5%)	4 (3%)	11 (6%)	52 (5%)		
Other groups[a]	27 (2%)	3 (1%)	3 (2%)	7 (4%)	14 (1%)		
Not reported	419 (25%)	39 (13%)	17 (14%)	57 (31%)	306 (29%)		
Primary language						−	−
English	1110 (67%)	184 (59%)	78 (64%)	122 (67%)	726 (70%)		
Spanish	313 (19%)	70 (23%)	31 (26%)	40 (22%)	172 (17%)		
Other	215 (13%)	53 (17%)	10 (8%)	21 (11%)	131 (13%)		
Not reported	16 (1%)	4 (1%)	2 (2%)	0 (0%)	10 (1%)		
Health Insurance						--	--
Public	1083 (60%)	284 (68%)	97 (62%)	116 (62%)	586 (56%)		
Private	570 (31%)	98 (23%)	35 (22%)	59 (32%)	378 (36%)		
Not reported	160 (9%)	38 (9%)	24 (15%)	12 (6%)	86 (8%)		
Stage 1 screen							
Child age (mean months)	23.8 (5.5)	24.6 (4.9)	26.3 (4.8)	22 (5.6)	23.4 (5.5)	*	−
BITSEA—ASD problem	2.3 (2.7)	5.7 (3.1)	2.6 (1.6)	3.1 (3)	1.2 (1.4)	**	**
BITSEA—ASD competence	12.1 (3)	8.9 (3.1)	11.5 (2.2)	10.8 (2.9)	13.3 (2.2)	**	**
Diagnostic evaluation							
Number completed	237 (14%)	178 (57%)	46 (38%)	13 (7%)	N/A		
Child age (months)	27.4 (5.5)	27.3 (4.9)	29.4 (4.8)	25.0 (4.9)	N/A	−	−
MSEL—expressive lang	26.8 (7.7)	25.9 (7.2)	29.8 (8.9)	27.8 (6.5)	N/A	**	−
MSEL—receptive lang	22.7 (7.5)	22.1 (6.9)	24.4 (8.6)	26.1 (9.9)	N/A	−	−
MSEL—early learning composite	60.6 (11.1)	59.2 (10.7)	63.6 (11.7)	70.8 (8.2)	N/A	*	**
Vineland—adaptive behavior composite	64.1 (10.0)	62.1 (9.9)	70.0 (7.9)	67.0 (4.5)	N/A	**	−
ADOS calibrated severity score	8.0 (2.2)	8.2 (2.1)	7.4 (2.3)	6.8 (2.0)	N/A	−	−

BITSEA brief infant toddler social emotional assessment, *ASD* autism spectrum disorder, *BITSEA* brief infant toddler social emotional scale, *MSEL* mullen scales of early learning, *ADOS* autism diagnostic interview schedule

[a]Other groups include: Jewish, Hebrew, Ashkenazi, & French Creole

Vineland Adaptive Behavior Inventory. Based on stage 1 results, 37% of children were referred to stage 2. Figure 1 presents a proportional Venn diagram that depicts the co-occurrence of each referral criterion: (1) stage 1 positive screening scores (present among 80% of children who were referred to stage 2); (2) EI provider concern (64% of children

who were referred to stage 2); and (3) EI provider report of parent concern (38% of children who were referred to stage 2). Overall, 50% of children who were referred to stage 2 had both a positive screening score and report of provider/parental concern (path A) 20% had concern only (path B), and 30% had only a positive screening score with neither provider nor parental concern (path C).

Figure 2 depicts paths A, B, and C from stage 1 screening to diagnosis in greater detail. Across levels of analyses, children in Path A continued to participate in the screening process and to receive confirmatory results at rates that were either equivalent to or higher than children in Path B, whose rates in turn were either equivalent to or higher than those for children in Path C. In particular, children in Path A and Path B were more likely to complete stage 2 screening than children in Path C ($\chi^2(1) = 91.4$, $p < 0.001$ and $\chi^2(1) = 46.2$, $p < 0.001$, respectively), while the difference between Paths A and B was not significant ($\chi^2(1) = 1.2$, $p = 0.28$). As a result of stage 2 screening, children in Path A were more likely to be referred for diagnostic assessments than children in Path B ($\chi^2(1) = 11.6$, $p < 0.001$), who in turn were more likely to be referred than children in Path C ($\chi^2(1) = 15.8$, $p < 0.001$). Children in Path A were more likely to complete diagnostic evaluations than children in Path B ($\chi^2(1) = 3.9$, $p < 0.05$) and Path C ($\chi^2(1) = 6.6$, $p < 0.01$), while the difference between Paths B and C were not significant ($\chi^2(1) = 1.1$, $p = 0.29$). Finally, children in Path A were more likely to receive an ASD diagnosis than were children in Paths B and C ($\chi^2(1) = 6.6$, $p < 0.01$ and
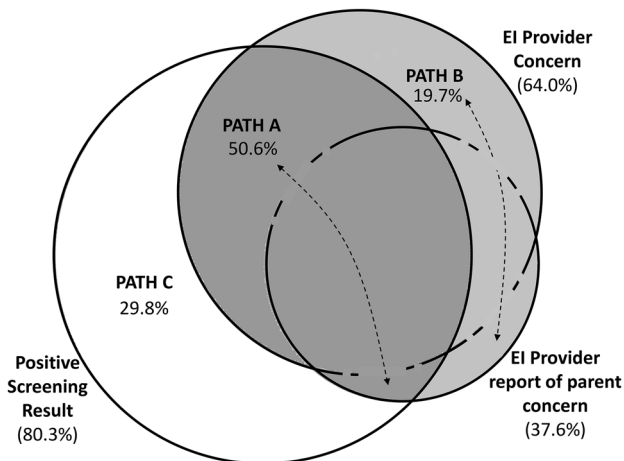

Fig. 1 For children referred to stage 2 (which is the denominator of this proportional Venn diagram), the circles depict the prevalence of each referral criterion: positive screening score (80.3%), EI provider concern (64.0%), and EI provider report of parent concern (37.6%). Intersections of the circles define the different pathways to referral: Path A (50.6%) = Positive screening result + concern; Path B (19.7%) = Concern only; Path C (29.8%) = Positive screening result only
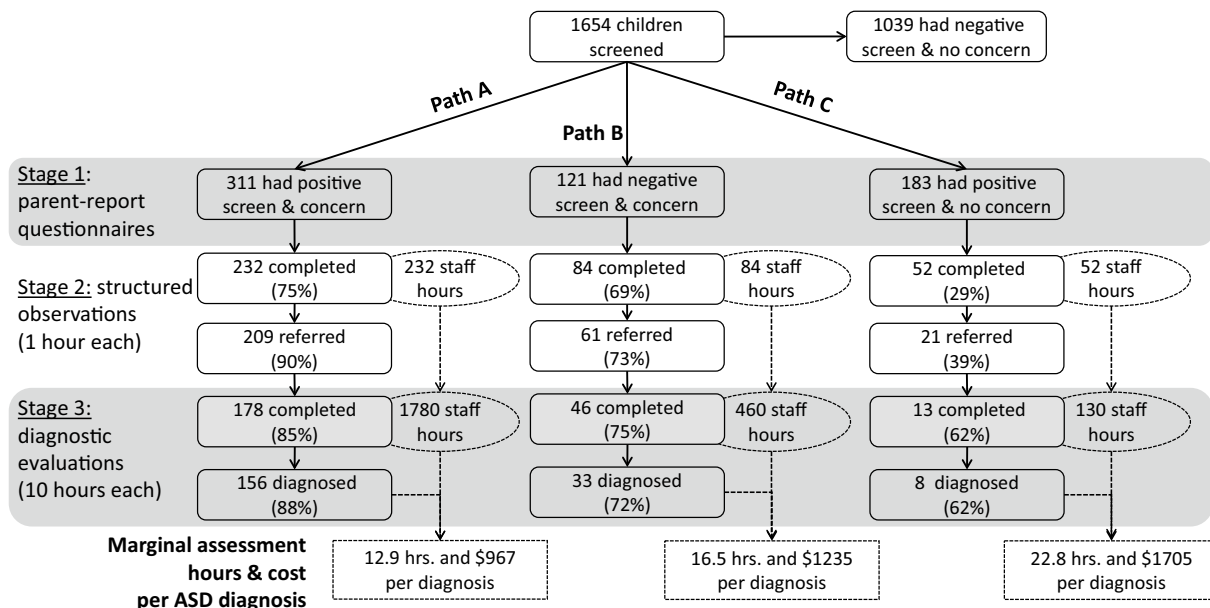

Fig. 2 For each referral pathway, boxes report the number and percentage of families to complete each stage of assessment and be referred to the next stage. Ellipses and dotted lines depict the total number of staff hours to complete each stage of assessment in each referral pathway. Cost per hour of clinical assessment was estimated based on Medicaid/Masshealth rates in 2017

$\chi^2(1) = 5.9$, $p < 0.05$, respectively), while the difference between Paths B and C was not significant ($\chi^2(1) = 0.5$, $p = 0.48$). Survival analyses yielded results that were consistent with these analyses. Time-to-completion of diagnostic assessment was superior in Path B compared to Path C ($\chi^2(1) = 7.4$, $p < 0.01$), but no difference was found between Paths A and B ($\chi^2(1) = 1.0$, $p = 0.32$). This result was consistent with comparable analyses conducted on raw data with case wise deletion, and no differences were found across EI sites.

Participant observations and video review revealed that stage 2 assessments required an average of 1 staff hour, while stage 3 diagnostic evaluations required 6 staff hours for administration and feedback, plus an additional 4 h for scoring and report writing. Overall, administration of stage 2 and stage 3 assessments for Path A required 2012 staff hours and yielded 156 ASD diagnoses, or 12.9 h and $967 per diagnosis. Administration of stage 2 and stage 3 assessments for Path B required 544 staff hours and yielded 33 ASD diagnoses, or 16.5 h and $1235 per diagnosis. Administration of stage 2 and stage 3 assessments for Path C required 182 staff hours and yielded 8 ASD diagnoses, or 22.8 h and $1705 per diagnosis. Overall, conducting follow-up evaluations for children in Path B proved to be more cost-effective in regard to case finding than for children in Path C. Because referring children based on positive screening results is widely recommended, this evidence suggests that it is also reasonable to refer children based on presence of concern -- even when screeners are not positive for ASD risk.

Aim 2: Quantitative results

Figure 3 describes the relationship between screening score, provider concerns, perception of parent concerns, and chance of completing stage 2 screening. The distribution of screening scores is skewed to the right in that the majority of children received relatively low scores, while a minority received higher scores. The probability of concern increased with screening score (as depicted in the shading of the bars). The lines depicting probabilities reflect two logistic regression analyses, one a strong effect of screening score on concern (OR = 1.4, $p < 0.001$), and a second showing a strong effect of screening score on completion of stage 2 screening (OR = 1.1, $p < 0.001$). While these analyses document strong associations, no evidence of a threshold effect was apparent (i.e., where small changes in screening results predicted large changes in concern or completion of stage 2), as would be expected if EI providers responded to positive and negative scores rather than overall results.
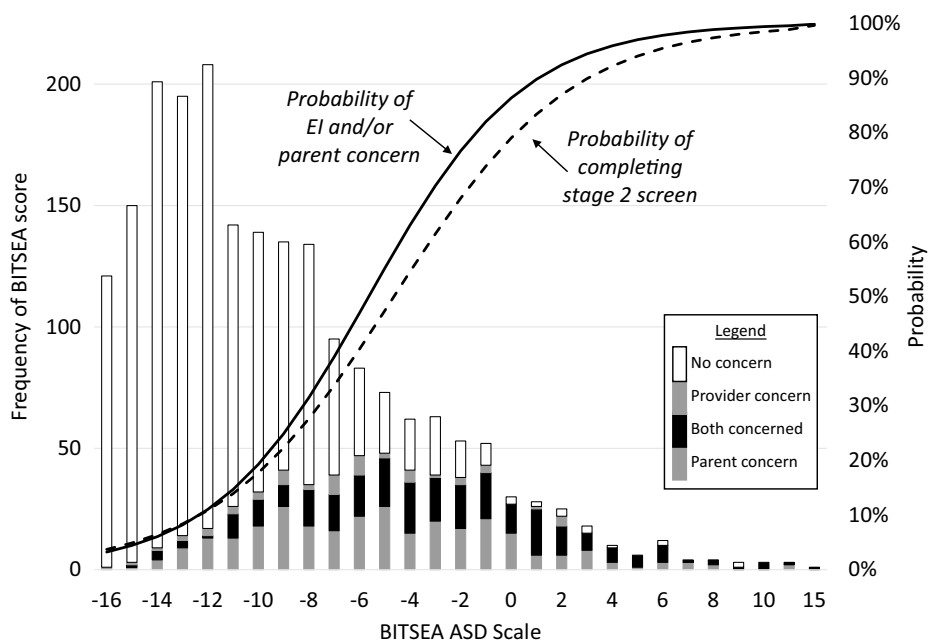

Fig. 3 BITSEA = Brief Infant Toddler Social Emotional Assessment. The histogram represents that frequency of scores on the BITSEA ASD Scale. For each score, stacked histogram values depict the relative frequency of different levels of concern (see legend). The solid line labeled "probability of EI and/or parent concern" reflects the results of a logistic regression predicting probability of concern based on screening score; probabilities therefore approximate the proportion of families for whom there was concern (shaded in the stacked histogram) divided by the total number of families to receive each screening score. Likewise, the dotted line labeled "probability of completing stage 2 screening" reflects the results of a logistic regression predicting probability of completing the stage 2 screen based on screening score

To evaluate the three variables included in our decision rule, we used a Cox regression model to conduct a semiparametric survival analysis of time-to-complete stage 2 screening subsequent to stage 1 screening. Independent variables included a positive screening result, EI provider concern, and EI provider perception of parent concern. In an initial model testing main effects, only the two concerns variables were statistically significant; positive screen was not. To evaluate the influence of interactions between independent variables, we tested another model that included three additional dummy variables indicating 2-way interactions among the three main effects. None of the interaction terms were statistically significant, and a likelihood ratio test revealed that the interaction model did not offer significantly better fit to the data than the main effects model ($\chi^2(3) = 1.6$, $p > 0.10$); therefore, the main effects model was retained for analysis. As displayed in Fig. 4a, expected values based on the Cox model fit observed values well for the seven combinations of IVs that resulted in referral. Post hoc Wald tests revealed that these seven pathways clustered into three groups, within which there were no significant differences. These groups were characterized by number of

concerns, i.e., two concerns (EI provider concern and perception of parent concern), 1 concern (EI provider concern or perception of parent concern), or no concern reported (positive screen only). Figure 4b presents a graph of completion rates (i.e., 1—survival) by time, as well as median time-to-completion among those who completed time 2 screening. Median times were inversely related to completion rates: half of cases with 2 concerns who completed stage 2 did so within 28 days, half of cases with 1 concern who completed stage 2 did so within 42 days, and half of cases with no concerns who completed did so within 55 days.

Aim 2: Qualitative Results

Qualitative analyses investigated the process of screening and shared decision-making, including the strategies deployed by the EI providers to adapt to the presence or absence of concern during the multi-stage screening protocol. Our findings suggest that the concerns expressed by providers and parents influenced both the initiation of and ongoing engagement with the multi-stage screening protocol. While providers attributed attrition in the multi-stage
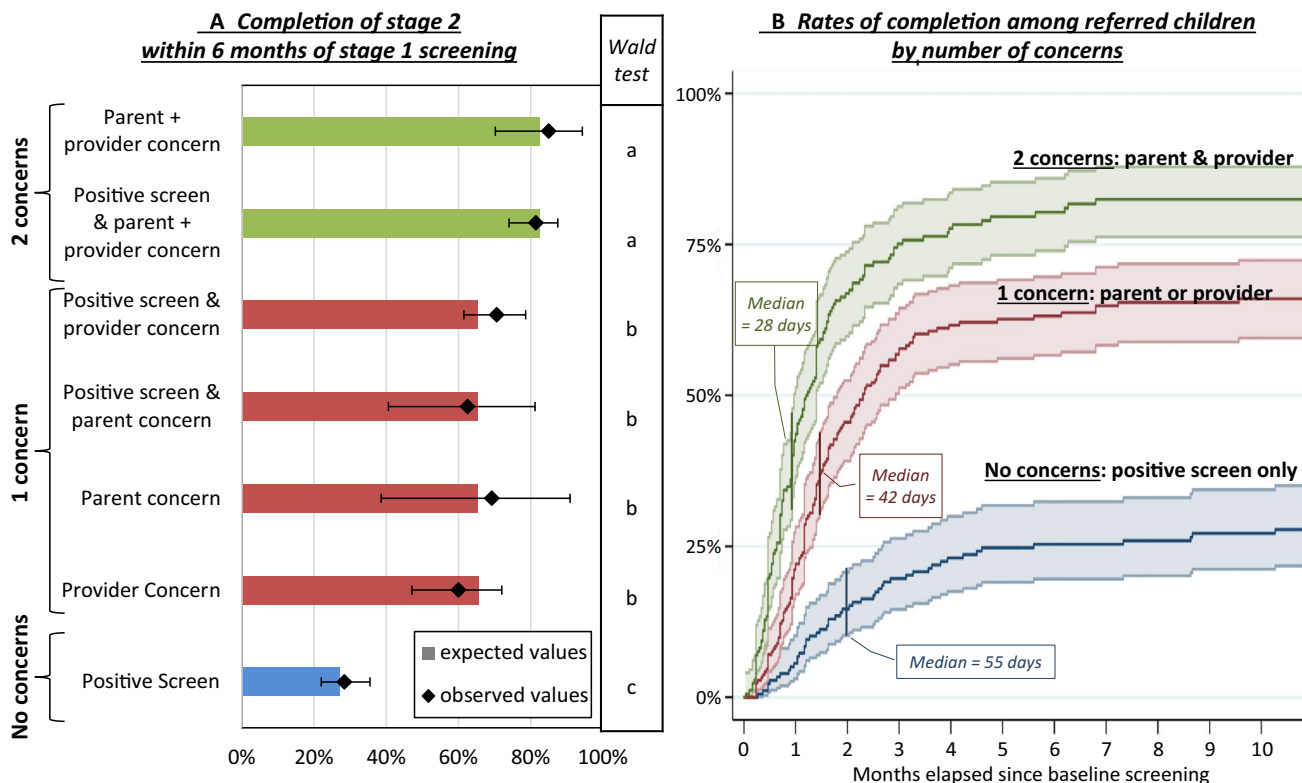

Fig. 4 For each combination of referral criteria, the left panel compares the proportion of families expected to complete stage 2 screening based on the survival analysis (as depicted in the histogram) to those observed to complete stage 2 screening (as depicted by the diamonds with 95% confidence intervals). Results of Wald tests,

depicted in the middle panel, reveal that families with two concerns completed stage 2 at higher rates than families with 1 concern, who in turn completed stage 2 at higher rates than families with 1 concern. The panel on the right depicts completion curves for each of these groups along with median days to completion

screening protocol to a range of causes that included their inability to schedule assessments (n = 4), families moving to another jurisdiction (n = 2), and children aging out of eligibility for EI services (n = 2), by far the most common reason cited for not continuing in the two stage process was a lack of parental concern and readiness (n = 10 of 21 respondents).

In cases where parents lacked concerns, the progress through the screening protocol was frequently delayed as families revisited their decisions over time, as illustrated below:

> …with one family, back in May, and the girl had a concerning score, but mom did not want to move forward…. Slowly [with each session] mom started talking to us a little bit more about the things she was seeing and those red flag behaviors [from the stage one screening tool], and then we just did the [stage two screener] a couple of weeks ago. [EI provider]

Notably, providers strategically articulated concerning behaviors "slowly with each session" [EI Provider] to help parents perceive the same concerns. Illustrative of this theme, one provider articulated the incremental process of sharing concerns around ASD:

> from the start, I've been dropping breadcrumbs, but I suddenly realize, like, 'okay, I need some really bigger crumbs to move us along. [EI provider]

As described in the full narrative, this provider introduced ASD concerns gradually to avoid provoking the family to 'drop out' early. Consistent with this finding, parents also remarked upon providers deliberately slowing the process down to accommodate their concerns. Parents frequently delineated the role of the EI provider to include making recommendations, rather than giving instructions, as illustrated in the following quote:

> "No, I think they were really patient with us, and they – deep down, they probably wanted us to get him screen earlier, but obviously they're not gonna say, "You should do this." Like, they're not gonna tell us straight out, "You should do this." They recommend. So, I feel like they probably wanted us to get him screened earlier, and I'm glad they were patient with us." [Parent]

Providers expressed that strategies to slow the process down were strongly influenced by the provider's commitment to an ongoing process of parental informed consent, as illustrated by the following quote:

> I make sure that they know at any point it is all optional. You know even if they go to the STAT (and) it's concerning they don't actually have to have a developmental [evaluation] done, but just because they're

saying yes to the first questionnaire they're not like locked into the process…one family just really was not a fan of it." [EI Provider]

Therefore, in cases where providers perceive a lack of parental concern, delays are more likely to emerge as demonstrated in findings of delay in time to diagnosis in Path C.

EI providers also emphasized limitations of relying on the screening result alone without presence of a clinical concern. As one provider states:

> I'd say [the screening tool] catches probably more kids than just ASD concerns. But, yeah, I think it does usually catch all of the kids that I would be concerned about and then some. [EI Provider]

Accordingly, most respondents indicated less confidence in the screening tool when results were not consistent with their own concerns.

At the same time, some providers indicated a positive screening tool sometimes elicited concern in cases where they had not previously held it. As illustrated below, one provider suggests that the screening tool assisted in their own identification of concern for ASD:

> The mom had been told before, actually, by another professional, and she never conveyed it to me, and it wasn't until that BITSEA-POSI came out positive that I was sort of like, oh my gosh, like, how did I miss this, sort of. So… and she, you know, and that started a whole conversation, and it was very telling to me, like, not only are we getting this positive screen, but now all of a sudden Mom's telling me that so-and-so said that her child might be exhibiting signs as well, so yeah, no, I think it's caught. I know for a fact, one child it's caught, that I was like, oh, I didn't even notice, so, yeah. [EI provider]

In summary, EI providers often discounted positive screening results that were inconsistent with their own perspectives and concerns, but sometimes positive screening elicited concerns that led to further disclosures and engagement with the multi-stage screening protocol.

Discussion

Consistent with previous evidence (Godoy and Carter 2013; Godoy et al. 2014), results of this study suggest that parents' concerns, providers' clinical judgment, and shared decision-making can be important drivers of the detection and diagnosis of ASD. Even in the context of an evidence-based screening protocol, parents' and providers' concerns were more predictive of referral completion than were positive scores on the screener. Our clinical decision rule that encouraged

referrals based on providers' and/or parents' concerns even in the absence of positive screening results performed well, yielding more ASD diagnoses per staff hour than referrals based on positive screening results in the absence of clinical concern. Results also suggest that EI providers respond not only to screening results, but also to parents' opinions and perspectives when guiding children through the multi-stage screening process. In short, presence of concerns facilitated initiation of and engagement with the screening protocol, while lack of concerns among parents induced providers to move more slowly, to support continued family engagement. Results thus present a case study in how the process of shared decision-making between parents and providers can drive the detection and diagnosis of ASD.

We note several limitations to our study. Perhaps most important are the inherent limits to generalizability. The utility of clinical judgment is very likely dependent on context, including the training of providers and the time, support and resources available for evaluation. While we argue that our results provide compelling evidence that inclusion of clinical judgment in our decision rule contributed to the success of our protocol, the utility of clinical judgment in other contexts should not be assumed but, instead, further evaluated. Moreover, the importance of shared decision-making emerged as a finding of our qualitative research but was only indirectly evaluated in our quantitative analysis. For example, assessment of providers' and parents' concerns each relied on single questions, both of which were completed by the EI provider. Future research on the role of clinical judgment would benefit from a more robust assessment of concerns among all decision-makers who have a stake in the choice at hand, as well as a more explicit focus on strategies for shared decision-making.

Our study was also limited in its lack of complete diagnostic information for children who screened negative at stages 1 or 2 or whose parents chose not to participate. Because the study focused on implementation science and not on diagnostic accuracy, we can only assess the utility, but not the accuracy, of initial screening results and reports of concern. Moreover, we were unable to fully assess the causes of discordance between screening results, parents' concerns, and providers' concerns, such as whether such disagreements are based on different perceptions of risk for ASD, or on different perceptions regarding the costs and benefits of being evaluated and potentially receiving a diagnosis. Differences in risk perception are likely for a number of reasons. On the one hand, parents and providers have access to a wide range of information that may not be directly reflected in screening scores, thus leading to more accurate perceptions of risk, but on the other hand they may also be prone to cognitive biases, thus reducing accuracy. Differences in risk tolerance are also possible. Even if families, providers, and screening instruments are in agreement

as to the severity of a child's symptoms and their risk for ASD, parents may have different perspectives on the utility of obtaining a diagnosis, perhaps because they have different perspectives on the costs and benefits and availability of treatments or from differing concerns about stigma. Further research is needed to disentangle the different possible explanations for why parents' and providers' concerns may be discordant with screening results. Moreover, in the absence of information regarding the diagnostic status of children whose parents chose to not pursue further screening or diagnostic assessment, it is impossible to consider the costs of missed early diagnoses (e.g., which may lead to higher special education costs).

Nevertheless, results have important implications for implementation science in regard to evidence-based screening. While the utility of clinical judgment cannot be assumed, it is equally true that the utility of clinical judgement should not be dismissed. When creating evidence-based protocols, implementers should carefully consider what decisions must be made, who is in the best position to make them, and how they should be made. We believe that our protocol was typical in emphasizing the administration of screening instruments as a mechanism to promote the goal of earlier detection and diagnosis. However, by promoting judgment among EI providers, we were able to increase the efficiency of our protocol to detect and diagnose cases of ASD.

In addition, whereas the goal of most screening protocols is to promote detection and diagnosis as early as possible, qualitative results suggest a second goal that is equally important to providers: continued family engagement. Failure to recognize the importance of maintaining family engagement through the screening and referral process risks creating a disconnect between front-line staff who provide care and researchers who implement evidence-based protocols. At best, such a disconnect could lead to unexplained adaptation to evidence-based protocols as providers deviate from protocol timelines to accommodate discordance between parental readiness, provider concerns, and screening results. Qualitative evidence for such adaptations in our study is consistent with observations that providers seldom refer all children who screen positive, as is commonly reported in the literature (Sheldrick et al. 2016). Moreover, the potential for such disconnects between parents, providers and screening results highlights the need for shared decision-making to promote and maintain family engagement. As recognized by the IOM/NAM, use of medical tests typically occurs within a context of shared decision-making (Balogh et al. 2015). This recognition offers potential targets for further research. Consistent with models of screening in the decision-making literature (Fryback and Thornbury 1991), respondents' emphasis on the importance of engagement and the role of concern suggests that the utility of screening

instruments may depend as much on their ability to influence perceptions and judgments (i.e., convincingness) as on their sensitivity and specificity (i.e., accuracy). Moreover, consistent with the IOM/NAM's call for more education to improve clinical judgment, these results speak to the importance of training EI providers on techniques to begin and maintain difficult conversations with parents, including addressing discordance between screening results and parent/provider concerns and using strategies such as motivational interviewing to align parents' and providers' goals (Gayes and Steele 2014). Screening protocols may be improved by developing an evidence base to support more detailed guidance for engaging in shared decision-making in regard to screening. Research on optimal strategies to respond to a perceived lack of parental readiness while addressing concerns about 'false positive' screening results would be particularly useful. Unfortunately, we know of little evidence that specifically focuses on effective communication with families regarding ASD screening and diagnosis. Addressing this evidence gap has the potential to improve efficiency by increasing family and provider engagement with screening.

Our study also highlights the utility of diverse methods for advancing implementation science. In particular, evidence regarding cost is often cited as important to decision-makers yet is under-represented in the research literature. Even the simple cost analysis we conducted was able to yield important insights about the relative cost-effectiveness of different observed referral pathways. Further analyses of costs from the perspective of different decision-makers would likely yield additional insight. Similarly, use of mixed methods guided by a decision-making framework proved to be useful in practice. In particular, focusing quantitative analysis on a choice regarding process (i.e., the decision rule for referrals) while focusing qualitative analyses on parents' and providers' decision-making yielded complementary results that suggest novel mechanisms for the effectiveness of screening that emphasize questionnaires' ability to promote communication and shared decision-making while also improving accuracy.

Finally, our study highlights both the importance and complexity of "concern" as a construct. On the one hand, parents' and providers' concerns were found to influence both initiation of and engagement with screening. On the other hand, their concerns were found to change over time in response to additional information, including the results of screening instruments. This complex and dynamic interaction suggests that screening and concern are not independent, but instead represent interdependent processes that interact to drive the detection and diagnosis of ASD.

Author Contributions RCS participated in the design of the study, conducted primary quantitative analyses, created figures, drafted sections of the manuscript, and directed the editing process. EF and JDV assisted with initial data analyses, writing of the initial draft, and editing of the final manuscript. TM conceptualized and directed all qualitative data analyses, wrote all qualitative sections, and edited the final manuscript. FMP conceptualized the overarching study design and edited the final manuscript. NH conducted descriptive data analyses, assisted with interpretation of results, and substantively edited the final manuscript. AE and AF participated in the design of the study, the interpretation of data, and the editing of the final manuscript. AS participated in the design of the study, conceptualization of the paper, interpretation of qualitative and quantitative evidence, and substantive revisions of the final manuscript.

Compliance with Ethical Standards

This research was supported in part by a NIMH grant to Drs. Carter and Sheldrick (R01MH104400). Dr. Sheldrick is the co-creator of the POSI, which is one of the two screeners used in this study. He conducts research related to this instrument but receives no royalties. Dr. Carter is the co-creator of the BITSEA, which is one of the two screeners used in this study. She receives royalties related to the licensing of this instrument.

Ethical Approval Ethical approval was granted by the Institutional Review Board of the University of Massachusetts Boston. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341–382. https://doi.org/10.1177/0011000005285875.

Balogh, E., Miller, B. T., Ball, J., & Institute of Medicine (U.S.). Committee on Diagnostic Error in Health Care. (2015). *Improving diagnosis in health care*. Washington, DC: The National Academies Press. https://doi.org/10.17226/21794.

Briggs-Gowan, M., & Carter, A. S. (2006). *Brief infant toddler social emotional assessment (BITSEA)*. (pp. 17–19) London: Pearson.

Brożek, J. L., Akl, E. A., Alonso-Coello, P., Lang, D., Jaeschke, R., Williams, J. W., … Schünemann, H. J. (2009, May). Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy: European Journal of Allergy and Clinical Immunology*. https://doi.org/10.1111/j.1398-9995.2009.01973.x.

Calzada, L. R., Pistrang, N., & Mandy, W. P. (2012). High-functioning autism and Asperger's disorder: Utility and meaning for families. *Journal of Autism and Developmental Disorders, 42*(2), 230–243.

Creswell, J. W., Klassen, A. C., Clark, P., V. L., & Smith, C., K (2011). *Best practices for mixed methods research in the health sciences.* Bethesda: Commissioned by the Office of Behavioral and Social Sciences Research.

Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care, 50*(3), 217.

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573.

Fryback, D. G., & Thornbury, J. R. (1991). The efficacy of diagnostic imaging. *Medical Decision Making, 11*(2), 88–94.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report, 20*(9), 1408–1416.

Gayes, L. A., & Steele, R. G. (2014). A meta-analysis of motivational interviewing interventions for pediatric health behavior change. *Journal of Consulting and Clinical Psychology, 82*(3), 521.

Giserman Kiss, I., Feldman, M. S., Sheldrick, R. C., & Carter, A. S. (2017). Developing autism screening criteria for the brief infant toddler social emotional assessment (BITSEA). *Journal of Autism and Developmental Disorders, 47*(5), 1269–1277. https://doi.org/10.1007/s10803-017-3044-1.

Godoy, L., & Carter, A. S. (2013). Identifying and addressing mental health risks and problems in primary care pediatric settings: A model to promote developmental and cultural competence. *American Journal of Orthopsychiatry, 83*(1), 73–88. https://doi.org/10.1111/ajop.12005.

Godoy, L., Mian, N. D., Eisenhower, A. S., & Carter, A. S. (2014). Pathways to service receipt: Modeling parent help-seeking for childhood mental health problems. *Administration and Policy in Mental Health and Mental Health Services Research, 41*(4), 469–479. https://doi.org/10.1007/s10488-013-0484-6.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293–323. https://doi.org/10.1037/1076-8971.2.2.293.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19.

Guinchat, V., Chamak, B., Bonniau, B., Bodeau, N., Perisse, D., Cohen, D., & Danion, A. (2012). Very early signs of autism reported by parents include many concerns not specific to autism criteria. *Research in Autism Spectrum Disorders, 6*(2), 589–601.

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215.

King, T. M., Tandon, S. D., Macias, M. M., et al. (2010). Implementing developmental screening and referrals: Lessons learned from a national project. *Pediatrics, 125*(2), 350–360.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). Autism diagnostic observation schedule (ADOS). *Journal of Autism and Developmental Disorders.* https://doi.org/10.1007/BF02211841.

McPheeters, M. L., Weitlauf, A., Vehorn, A., Taylor, C., Sathe, N. A., Krishnaswami, S., et al. (2015). Screening for autism spectrum disorder in young children: A systematic evidence review for the U.S. preventive services task force. *AHRQ Publication No., 13-05185, EF-1*, (121), 202.

Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee &

Medical Home Initiatives for Children With Special Needs Project Advisory Committee. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 118*(1), 405–420. https://doi.org/10.1542/peds.2006-1231.

Meehl, P. E. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology, 6*(2), 102–109. https://doi.org/10.1037/h0049190.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*(3), 194.

Mullen, E. M. (1995). Mullen scales of early learning, AGS Edition: Manual and item administrative books. (pp. 1–92). Noida: American Guidance Services Inc.

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., … Zwaigenbaum, L. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A baby siblings research consortium study. *Journal of Child Psychology and Psychiatry, 56*(9), 988–998.

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine, 302*(20), 1109–1117.

Pizur-Barnekow, K., Muusz, M., McKenna, C., O'Connor, E., & Cutler, A. (2013). Service coordinators' perceptions of autism-specific screening and referral practices in early intervention. *Topics in Early Childhood Special Education, 33*(3), 153–161.

Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T., & Fein, D. (2013). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics, 133*(1), 37–45.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*(2), 131–144.

Salisbury, L., Nyce, J. D., Hannum, C., Sheldrick, R. C., & Perrin, E. C. (2018). Sensitivity and Specificity of 2 Autism Screeners Among Referred Children Between 16 and 48 Months of Age. *Journal of Developmental and Behavioral Pediatrics, 39*, 254–258.

Sanders, S., Doust, J., & Glasziou, P. (2015). A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. *PLOS ONE, 10*(6), e0128233. https://doi.org/10.1371/journal.pone.0128233.

Sheldrick, R. C., Benneyan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Thresholds and accuracy in screening tools for early detection of psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 56*(9), 936–948. https://doi.org/10.1111/jcpp.12442.

Sheldrick, R. C., Breuer, D. J., Hassan, R., Chan, K., Polk, D., & Benneyan, J. (2016). A system dynamics model of clinical decision thresholds for the detection of developmental-behavioral disorders. *Implementation Science, 11*, 1–14. https://doi.org/10.1186/s13012-016-0517-0.

Sheldrick, R. C., & Garfinkel, D. (2017). Is a Positive developmental-behavioral screening score sufficient to justify referral? A review of evidence and theory. *Academic Pediatrics, 17*(5), 464–470. https://doi.org/10.1016/j.acap.2017.01.016.

Smith, N. J., Sheldrick, R. C., & Perrin, E. C. (2013). An abbreviated screening instrument for autism spectrum disorders. *Infant Mental Health Journal, 34*(2), 149–155. https://doi.org/10.1002/imhj.21356.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland adaptive behavior scales*, (3rd ed.). London: Pearson.

Stahmer, A. C. (2007). The basic structure of community early intervention programs for children with autism: Provider descriptions.

Journal of Autism and Developmental Disorders, 37(7), 1344–1354.

Stahmer, A. C., Collings, N. M., & Palinkas, L. A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities, 20*(2), 66–79.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: Screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders, 30*(6), 607–612. https://doi.org/10.1023/A:1005647629002.

Stone, W. L., McMahon, C. R., & Henderson, L. M. (2008). Use of the screening tool for autism in two-year-olds (STAT) for children under 24 months. *Autism, 12*(5), 557–573. https://doi.org/10.1177/1362361308096403.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1–26. https://doi.org/10.1111/1529-1006.001.

Tobin, G. A., & Begley, C. M. (2004). Methodological rigour within a qualitative framework. *Journal of Advanced Nursing, 48*(4), 388–396.

Trikalinos, T. A., Siebert, U., & Lau, J. (2009). Decision-analytic modeling to evaluate benefits and harms of medical tests: Uses and limitations. *Medical Decision Making, 29*(5), E22–E29. https://doi.org/10.1177/0272989X09345022.

Wiggins, L. D., Levy, S. E., Daniels, J., Schieve, L., Croen, L. A., DiGuiseppi, C., et al. (2015). Autism spectrum disorder symptoms among children enrolled in the Study to Explore Early Development (SEED). *Journal of Autism and Developmental Disorders, 45*(10), 3183–3194.

Willms, D. G., Best, J. A., Taylor, D. W., Gilbert, J. R., Wilson, D. M. C., Lindsay, E. A., & Singer, J. (1990). A Systematic approach for using qualitative methods in primary prevention research. *Medical Anthropology Quarterly, 4*(4), 391–409. https://doi.org/10.2307/649223.

Zuckerman, K. E., Sinche, B., Cobian, M., Cervantes, M., Mejia, A., Becker, T., & Nicolaidis, C. (2014a). Conceptualization of autism in the Latino community and its relationship with early diagnosis. *Journal of Developmental and Behavioral Pediatrics: JDBP, 35*(8), 522.

Zuckerman, K. E., Sinche, B., Mejia, A., Cobian, M., Becker, T., & Nicolaidis, C. (2014b). Latino parents' perspectives on barriers to autism diagnosis. *Academic Pediatrics, 14*(3), 301–308.

Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., et al. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics, 136*(Supplement 1), S41–S59.