6-30-2016

Use of M-CHAT-R/F-A to Screen for ASD in Albania

Laura A. Brennan
University of Connecticut - Storrs, laura.brennan@uconn.edu

Use of M-CHAT-R/F-A to Screen for ASD in Albania

Laura Brennan, Ph.D.

University of Connecticut, 2016

The current study examined the screening and evaluation of 2,594 toddlers between 16 and 36 months of age for Autism Spectrum Disorder (ASD) to determine the clinical utility of Modified Checklist for Autism in Toddlers-Albanian version(M-CHAT-R-A). Of these children, 253 (9.75%) toddlers screened positive on the initial screening at pediatric well child visits in Albanian speaking communities in Albania. Follow up interviews were conducted by phone on failed items, as on the English version of the M-CHAT-R/F. Interviews were completed with 127 (50%) of the screened positive toddlers; 50% (n=126) of the 253 screened positive toddlers were lost to follow-up due to parent refusal or inability to contact. A total of 26 toddlers (21%) continued to screen positive after follow-up, qualifying them for a free developmental and diagnostic evaluation, 20 of whom received an evaluation. Autism Diagnostic Observation Schedule (ADOS) results indicated that 16% (n=3) of toddlers met criteria for ASD, 74% (n=14) met Autism criteria, and 11% (n=2) were classified as Non-ASD. Positive Predictive Value (PPV) of the 2-Stage M-CHAT-R-A for ASD or Autism was 0.895 (95% CI: 0.65-0.98) and 1.00 (95% CI: 0.79-1.00) for other developmental delays, with no typical children found in the evaluated sample. Results indicated that removal of three items, using a cut-off of 5 failed items on initial screening, and a cut-off of 2 failed items on follow-up interview would increase reliability and decrease false positive rates. Further research is needed to confirm the performance of this version of the screener and proposed cut-offs. Results also suggest that key presenting features of autism are comparable between the original population on which the M-CHAT and M-CHAT-R were developed and the Albanian population, despite difference in culture, language, and location.

Use of M-CHAT-R/F-A to Screen for ASD in Albania

Laura Anne Brennan

B.A., College of the Holy Cross, 2004
M.A., University of Connecticut, 2013

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2016

APPROVAL PAGE

Doctor of Philosophy Dissertation

Use of M-CHAT-R/F-A to Screen for ASD in Albania

Presented by

Laura Anne Brennan, B.A., M.A.

Major Advisor

_____

Marianne Barton, Ph.D.

Associate Advisor

_____

Deborah Fein, Ph.D.

Associate Advisor

_____

Jeffrey Fisher, Ph.D.

University of Connecticut
2016

ACKNOWLEDGEMENTS

I would like to first thank the children, families, and pediatricians in Albania for their time and participation in this project. To the Albanian Children's Foundation (ACF) and our study collaborators both in Albania and the U.K., especially Drs. Ariel Como and Iris Rathwell, I would like to express my admiration for your dedication to the project and to children with ASD in Albania. Your efforts across the many initiatives within this project have changed the landscape of service provision for children with ASD tremendously. I am particularly thankful to Ms. Erjona Capari for your incredible attention to detail and commitment to this study. I would also like to convey my sincere appreciation to Autism Speaks for creating a visionary initiative that funds projects such as this.

To my advisors, Marianne Barton, Ph.D. and Deborah Fein, Ph.D., my gratitude and respect for you both cannot be overstated. The passion and command you bring to the research and practice of expanding mental health services to all children and families, and to training clinical psychologists, is nothing short of inspiring. Thank you for the knowledge, mentorship, and support you have given me throughout my graduate training. Thank you also to Chi-Ming Chen, Ph.D. for your guidance on statistical and research design and your investment in my learning. I thank Jeffery Fisher, Ph.D. for his willingness to serve on my committee and provide valuable feedback on this topic, and to Marie Coppola, Ph.D. for her eagerness to offer guidance in this process as well.

Finally, thank you to my husband and son. Your patience, encouragement, and humor made achieving this momentous goal possible – and all the more sweet.

TABLE OF CONTENTS

Use of M-CHAT-R/F-A to Screen for ASD in Albania

Introduction

Significant disparities exist in the accessibility and quality of mental health services provided across the world (Saxena, Thornicroft, Knapp, & Whiteford, 2007). Currently, the availability of mental health resources depends largely on a country's level of economic development, with persons from lower income countries having significantly less access to mental health services than those from wealthier countries (Patel, Kieling, Maulik, & Divan, 2013). Levels of national economic development can be broadly parsed using the World Bank's classifications based on gross national income (GNI) per capita, designations which consist of low income, middle-income (separated into lower-middle income and upper-middle income), and high income economies (The World Bank, 2015; see Table 1 for GNIs corresponding with each income level). Individuals from lower- and middle-income countries (LMICs) have been shown to have fewer mental health services than those from high-income countries (HICs).

Young people from LMICs are of particular concern, as almost one third of the world's population is comprised of children and adolescents, 90% of whom live in LMICs (Kieling, Baker-Henningham, Belfer, Conti, Ertem, Omigbodun, et al., 2011). With recent evidence linking the global burden of disease in children under 10 years of age largely to developmental disabilities and emotional and disruptive behavior disorders, the need for mental health services in this population is a significant public health concern (Patel et al., 2013). Global burden of disease estimates also indicate that mental health disorders in childhood and adolescence account for 10-20% of the disability-adjusted life-years (DALYs) lost (Kieling et al., 2011). DALYs are used as an index of the global burden of disease for mental health problems in children and adolescents, by quantifying the number of years of "'healthy' life" lost as a result of a person's

disorder (World Health Organization, n.d.) The World Health Organization (WHO) uses this metric to understand discrepancies between the current state of a region's public health and the ideal, in which all persons live to advanced age, unhindered by disease or disability (WHO, n.d.). Mental health disorders have been found to contribute to as high as 30% of DALYs in young people, making access to high-quality mental health services even more urgent. Despite a growing body of research highlighting the need for efficacious and low-cost mental health services for children in LMICs and efforts by the WHO to create guidelines for mental health services in these areas, mental health resources are not yet available to a majority of children in LMICs (Patel, Chowdhary, Rahman, & Verdeli, 2011; Patel, Flisher, Nikapota, & Malhotra, 2008; World Health Organization, 2010). Autism Spectrum Disorder (ASD) is one of many behavioral disorders for which there is limited availability of diagnostic and treatment services.

ASD is a neurodevelopmental disorder that significantly impacts child functioning across both HICs and LMICs (Whiteford, Degenhardt, Rehm, Baxter, Ferrari, Erskine, ... & Burstein, 2013). ASD is characterized by a set of behavioral characteristics to include difficulties in social communication, social reciprocity, and interest in peers, as well as demonstration of restricted, repetitive behaviors or interests (American Psychiatric Association, 2013). The ASD global burden of disease is estimated to account for 4.2% (3.2-5.3%) of all DALYs (Whiteford et al., 2013). While prevalence rates are well-established in North America and Europe, studies investigating the prevalence of ASD in LMICs are fewer and vary dramatically in methodological approach (Elsabbagh, Divan, Koh, Kim, Kauchali, Marcín et al., 2012; Patel et al., 2008; Wallace, Fein, Rosanoff, Dawson, Hossain, Brennan et al.., 2012). Current U.S. estimates suggest that 1 out of 68 children carry an ASD diagnosis (Center for Disease Control, 2014). Recent global ASD prevalence estimates indicate 62 out of 10,000 children meet criteria

for ASD (Elsabbagh et al., 2012). However, large regions such as South East Asia, Africa, and eastern areas of the Mediterranean were not represented in this estimate, suggesting that the figure cannot be interpreted as more than a rough approximation (Elsabbagh et al., 2012). Limited diagnostic resources in LMICs, including limited professional and lay knowledge of ASD and limited access to reliable, valid, and culturally appropriate screening and assessment tools, often restrict researchers' ability to conduct epidemiological studies such as those included in this estimate of global ASD prevalence (Soto, Linas, Jacobstein, Biel, Migdal, & Anthony, 2014; Elsabbagh et al., 2012; Wallace et al., 2012). Further, where these resources are more accessible, as in areas of Western and Eastern Europe, study design factors including age of diagnosis, diagnostic criteria, and location also contribute significant variability, impacting the accuracy of prevalence rate estimates (Posada, Primo, Ferrari, & Martín-arribas, 2007; Williams, Higgins, & Brayne, 2006).

Despite the limitations of ASD research in LMICs, recent developments within the field indicate a shift toward prioritizing such work. The number of "ASD-specific" organizations established and studies published in LMICs have increased (Daley, Singhal, & Krishnamurthy, 2013). The WHO convened a summit in September 2013 to "develop a common agenda for action" to address the need for services, continued research, and collaboration across nations in order to meet the needs of those with ASD (WHO, 2013) Further, Autism Speaks, an organization originally dedicated to raising awareness of and improving treatments for children with ASD within North America and Western Europe, created the Global Autism Public Health (GAPH) Initiative to promote international awareness, research, and collaboration to benefit underserved populations with ASD across the world. GAPH currently cites "active partnerships, programs, or exploratory activities" across 42 countries (Autism Speaks, n.d.; Ashwood,

Buitelaar, Murphy, Spooren, & Charman, 2014). The current study is a result of one such program in Albania aimed at increasing awareness of ASD in the region, improving early detection and diagnosis of ASD, and establishing a rigorous intervention program to include training therapists in applied behavior analysis and developing video materials for a parent training program. When this large-scale project was originally conceived, Albania was considered by the World Bank to be a lower middle-income country (The World Bank, 2012). However, in 2013, Albania's GNI per capita increased and was subsequently re-classified as an upper middle-income country (The World Bank, 2013). Despite this increase in GNI, Albania remains one of the poorest European countries (USAID, 2011), and the services available to parents of children with ASD outside of this project remained limited.

Statement of Problem

There is a clear need for low-cost, effective measures for detecting and diagnosing ASD in children living in LMICs. It is now well documented that ASDs can be reliably diagnosed in children aged 18-24 months (Kleinman, Ventola, Pandey, Verbalis, Barton, Hodgson ... & Fein, 2008) and even earlier (Zweigenbaum, Bryson, Brian, Smith, Roberts, Szatmari, Roncadin, Garon, & Vaillancourt, 2015), and that early identification and intervention are associated with more positive outcomes (Rogers, 1996, Johnson & Myers, 2007; Robins and Dumont-Mathieu 2006; Anderson, Liang, & Lord, 2014). The use of parent-report screening instruments is widely regarded as a first step in detecting young children at risk for ASD (Soto et al., 2014; National Research Council (US) Committee on Educational Interventions for Children with Autism, 2001; Szatmari, Bryson, Boyle, et al., 2003). In the United States (U.S.), the American Academy of Pediatrics recommends utilizing an ASD-specific screener at 18- and 24-month well-child care (WCC) visits to screen for ASD symptoms (Johnson & Myers, 2007; Robins, Casagrande,

Barton, Chen, Dumont-Mathieu, & Fein, 2014). Theoretically, WCC screenings for ASD provide an opportunity for children who are integrated in the health care system and regularly attend WCC visits to be identified and receive services as early as possible.

However, current evidence in the U.S. suggests that median age at diagnosis continues to be over the age of 4 years, suggesting that children are not reliably identified within the recommended timeframe of 18-24 months (Center for Disease Control, 2014). Further, age of diagnosis in the U.S. is shown to be later in children who are non-white or from lower socioeconomic backgrounds (Mandell, Listerud, Levy, & Pinto-Martin, 2002; Robins et al., 2014). A recent study investigating the socioeconomic and racial disparities in ASD screening rates in the U.S. found that utilizing a WCC screening approach followed by timely diagnostic evaluations might minimize disparities in age at diagnosis (Herlihy et al., 2014). This literature suggests that even within HICs, disparities continue to exist and impact individuals' access to needed services.

Data regarding ASD prevalence and identification as well as information about the availability of services are less readily available across the European Union (EU). The EU's European Autism Information System (EU, EAIS) was established in response to the EU's recognition of this issue within its member states (Posada, Primo, Ferrari, & Martín-arribas, 2007). This organization developed and disseminated a questionnaire called the "Autism Spectrum Disorders Prevalence Data and Accessibility to Services Questionnaire (Q-EAIS)" to further the EUs understanding of this population (Posada et al., 2007). Seven countries and four regional areas completed the Q-EAIS. Findings from this report indicated that individuals with ASD were primarily identified within the public health care system, which was universal for residents in all but one of these countries (Posada et al., 2007). WCC visits at 18 months were

reported to be established in 45% of respondents, with 54% of the countries reporting that ASD early screening procedures were included in public health systems (Posada et al., 2007). Age at diagnosis differed by DSM-IV-TR ASD diagnostic category, with 72% of children diagnosed with Autistic Disorder between 3 and 4 years of age (Posada et al., 2007). Across the countries and regions participating in the survey, the degree to which procedures for early detection and diagnosis of ASD are outlined by health care systems varied greatly. Several countries did not have a standardized procedure for early detection, while others had well-defined processes (Posada et al., 2007).

Large-scale studies of ASD prevalence outside of the U.S. and Europe have primarily been conducted in Japan, South Korean, and China, with cited estimates of ASD in these countries being 1.81%, 2.64%, and 0.16%, respectively (Elsabbagh et al., 2012; Kawamura, Takahashi, & Ishii, 2008; Wong & Hui, 2008; Kim, Leventhal, Koh, Fombonne, Laska, Lim, ... & Song, 2011). A review of global ASD prevalence conducted by Elsabbagh and colleagues (2012) indicated that the bulk of ASD research has been carried out in high-income countries, apart from China, with a limited base of research arising from middle-income countries and no prevalence estimates found from low-income countries (Elsabbagh et al., 2012).

ASD Screening Tools

A number of validated measures are available to providers screening for ASD risk in children 18 months of age and older, including the Checklist for Autism in Toddlers (CHAT), the Modified Checklist for Autism in Toddlers (M-CHAT), and the Infant Toddler Checklist (ITC; Johnson and Meyers, 2007; Zweigenbaum, Bryson, Lord, Rogers, Carter, Carver, ... & Yirmiya, 2009). A recently validated, revised version of the M-CHAT, the M-CHAT-R with Follow up (M-CHAT-R/F) has demonstrated strong psychometric properties, with estimates of sensitivity

and specificity found to be 0.854 and 0.993, respectively (Robins et al., 2014). The M-CHAT-R/F is an amended version of the original M-CHAT, a 23-item parent-report measure designed to detect ASD symptoms in toddlers between the ages of 18 and 30 months (Robins, Fein, Barton, & Green, 2001; Chlebowski, Robins, Barton, & Fein, 2013). The M-CHAT itself is an adaptation of the Checklist for Autism in Toddlers (CHAT; Baron-Cohen, Allen, & Gillberg, 1992) that enables the CHAT to be used as a parent-report measure for ASD symptoms.

The M-CHAT-R/F relies on a two-stage screening process to assess risk for ASD. Stage 1 of screening consists of the M-CHAT-R, a paper and pencil measure involving 20 yes/no questions for parents to answer. The measure takes approximately five minutes to complete (Robins et al., 2014). The screener excludes three items previously on the M-CHAT screening tool, with additional changes including items reordered to reduce agreement bias, amended wording, and additional examples provided (Robins et al., 2014). Stage 2 of screening, the M-CHAT-R/F (follow up), involves a structured follow-up interview over the telephone or in person in which parents of screened positive children are asked questions about items failed and examples of concerning behaviors are solicited (Robins et al., 2014). This two-stage screening process yielded a positive predictive value (PPV) for receiving an ASD diagnosis of 0.475 and a PPV of 0.946 for any developmental delay or concern (Robins et al., 2014).

The 20 items on the M-CHAT-R screener (Stage 1) were organized to decrease agreement bias; three of these items were reverse scored, meaning these items were intended to be answered "no" as the 'passing' response, rather than "yes," for example, "have you ever wondered if your child is deaf?" (Robins et al., 2014). Three "foil" items pertaining to motor skills were also included to allow parents of delayed toddlers to answer "yes" to items on the screener. A child is considered to "screen positive" on the M-CHAT-R/F if he or she fails three or more items on the

screener during the first stage of screening. The child continues to screen positive if he or she fail two or more items on the follow-up interview during Stage 2. In the original validation studies of the M-CHAT and M-CHAT-R, children who screened positive for both screening stages received a free developmental and diagnostic evaluation through the study in order to establish diagnosis.

Use of Translated and Culturally Adapted Versions of the M-CHAT and M-CHAT-R/F

Due to its recent development, no published studies to date have validated the M-CHAT-R/F outside of the United States, though it has been translated for use in at least 21 languages. Its predecessor, the M-CHAT, however, has a demonstrated international presence. It has been translated into at least 69 languages and has been found to have clinical utility in a number of international validation studies, with sensitivity and specificity estimates ranging from .75 to 1 and .75 to .98 respectively, and PPVs ranging between .107 and .88 (Barton, Dumont-Mathieu, & Fein, 2011; Canal-Bedia, García-Primo, Martín-Cilleros, Santos-Borbujo, Guisuraga-Fernández, Herráez-García et al., 2010; Inada, Tomonori Koyama, Inokuchi, Kuroda, & Kamio, 2010; Kara, Mukaddes, Altinkaya, Guntepe, Gokcay, & Ozmen, 2014; Koh, Lim, Chan, Lin, Lim, Choo, & Magiati, 2013; Perera, Wijewardena, & Aluthwelage, 2009; Seif Eldin, Habib, Noufal, Farrag, Bazaid, Al-Sharbati, et al., 2008). Studies in Japan, China, and Singapore have also completed item-level analyses of the M-CHAT, two of which suggested that particular items better discriminated between groups of children with and without ASD in their samples (Inada et al., 2010; Koh et al., 2013; Wong, 2004). Items involving imitation, pretend play, pointing, showing, and social referencing were some of those suggested to best differentiate between diagnostic groups in these studies.

However, few international M-CHAT studies were able to complete the two-stage screening process utilized by the original M-CHAT and M-CHAT-R/F authors (Canal-Bedia et al., 2010; Kara et al., 2014; Seung, Ji, Kim, Sung, Youn, Hong, ... & Youm, 2015). A recent study used a translated, modified version (e.g., wording changes following pilot study) of the M-CHAT to screen for ASD in South Korea with the Korean M-CHAT, Version 2 (K-M-CHAT-2; Seung et al., 2015). The study examined screen positive rates at both stages of the screening process to evaluate the performance of the K-M-CHAT-2. Since formal diagnostic evaluations of screen positive children were not conducted, this study could not address the utility of the screening tool in identifying children with autism in Korea (Seung et al., 2015). In terms of screening results, however, this study demonstrated higher screen positive rates prior to follow up compared to the original M-CHAT study (26.4% in South Korea vs. 7.4% in the U.S.), while screen positive rates after follow up were comparable to those in the U.S. (2.3% and 3.0%, respectively; Seung et al., 2015, Robins et al., 2011).

During the first stage of a study investigating the utility of the M-CHAT in Turkey, investigators conducted Stage 2 follow up interviews with screened-positive toddlers; however, after determining a high false-positive rate, study procedures were modified to better suit the Turkish context, by training health professionals to verbally administer the M-CHAT to parents during pediatrician visits (Kara et al., 2014).

A study conducted in Spain most closely followed the original M-CHAT study using the two-stage screening procedures, with approval by the original M-CHAT study authors for the back-translated version of the Stage 2 questions (Canal-Bedia et al., 2010). This study suggested strong sensitivity and specificity for the M-CHAT in low- and high-risk samples in Spain, with a sensitivity of 1, specificity of .98, PPV of 0.35, and negative predictive value (NPV) of 1 (Canal-

Bedia et al., 2010). The authors suggest that lower prevalence rates (1 in 108 and 1 in 300) revealed by their study when compared to the U.S. may account for lower PPVs (Canal-Bedia et al., 2010). This study reported that Spanish M-CHAT adapted for use in Spain almost perfectly delineated between ASD and typically developing children, though the differences between ASD and those who received developmental delay diagnoses were less clear (Canal-Bedia et al., 2010). Results from this study also indicated that a cut-off score of five items failed or greater, rather than the original three or more failed, might reduce false positive rates in these samples. Although the effect of that change on sensitivity is unclear, it would likely reduce sensitivity to some extent. Alternative cut-off scores have been demonstrated by several international validation studies of the M-CHAT to better differentiate between toddlers with and without ASD in a particular country setting (Koh et al., 2013; Inada et al., 2011; Wong et al., 204; Kamio, Haraguchi, Stickley, Ogino, Ishitobi, & Takahashi, 2015).

Limitations of translated screening and diagnosis methods

As the fields of public health and ASD research evolve to incorporate a more global perspective on the need for universal access to quality mental health resources, the importance of conducting ASD screening research in a culturally appropriate manner and the understanding of how to do so has become an increasing focus of inquiry. Researchers from the U.S. and other high-income countries have begun to use translated versions of screening and diagnostic tools validated in their home nations as an attempt to address the limited understanding of ASD in middle- and low-income countries. This affords researchers the opportunity to broaden current conceptualizations of the disorder and preserves resources within LMICs by providing a measure already shown to be useful, rather than developing a measure independently (Daley, Singhal, & Krishnamurthy, 2013; Soto et al., 2015). However, several factors have been found to impact the

performance of such tools when used outside of the cultural and/or economic context in which they were validated. These include limited psychometric evidence regarding the utility of screening measures in middle- and low-income countries, assumptions that the nature of ASD symptoms are "universal" across these countries, and a limited appreciation of the "within-culture" factors such as socioeconomic status (SES, literacy rates, etc.) that might limit the effectiveness of screeners outside of high-income areas (Soto et al., 2015; Daley et al., 2013; Matson, Worley, Fodstad, Chung, Suh, Jhin, ... & Furniss, 2011). The impact of stigma, including the potential negative repercussions of receiving an ASD diagnosis in LMICs, is consistently a concern raised by researchers in this area (Soto et al., 2015; Daley et al., 2013).

A recent review of studies by Soto and colleagues (2015) focused on culturally adapted ASD screening tools and noted that few studies described in detail the process by which they achieved cultural adaptation. The authors rated each study based on the degree to which it followed the recommended process of translation, back-translation, committee review, and pilot testing (Soto et al., 2015). They concluded that the more rigorous the authors' approach to adaptation, the more likely their ASD screening tool was modified to best suit the context for which they were adapted. The changes these authors describe include wording alterations, inclusion of more culturally appropriate information or examples, procedural revisions, or alterative scoring criteria (Soto et al., 2015).

The current study is the first international study to utilize the original two-stage screening procedures outlined by the M-CHAT-R/F authors in order to determine its clinical utility as a screening tool outside of the U.S. This study utilized a translated version of the M-CHAT-R/F, the M-CHAT-R/F, Albanian Version (M-CHAT-R/F-A), administered to parents in Albania as

part of routine WCC visits. Before proceeding to a detailed review of the screening process, we provide brief background information about Albania.

Background: The Republic of Albania

The Republic of Albania lies in the south-eastern region of Europe on the Balkan Peninsula (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). Formerly a communist nation, Albania currently functions as a parliamentary democracy and has held three presidential elections (European Forum for Democracy and Solidarity, 2014; Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). According to the WHO, the current population of Albania is 3,173,000, with over 98% identified as ethnic Albanian (WHO, 2015; Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). The remaining 2% of the population in Albania identifies as Greek, Macedonian, Vlach, Roma, Bulgarian, or Serbian (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). A demographic report of the country published by Albania's Institutes of Statistics and of Public Health indicated that Albanians do not typically practice religion, but identify as Muslim, Catholic, or Orthodox Christian (Institute of Statistics & Institute of Public Health [Albania] and ICF Macro, 2010). The WHO reported that 54% of Albanians live in urban centers (WHO, 2015). The primary language in Albania is Albanian, largely consisting of two dialects, Gheg and Tosk (Simmons and Slocum, 2014). Of the two dialects, most modern Albanian establishments use Tosk, including its government, education system, media, and literature (Simmons and Slocum, 2014). According to the United Nations International Children's Emergency Fund (UNICEF) data, Albania has an adult literacy rate of 96.8% (UNICEF, 2013).

As stated previously, Albania is currently a middle-income country, with a GNI per capita of $4,460 (World Bank, 2014). In 2011, when Albania was considered lower-middle income, the

GNI per capita was $4, 390 (World Bank, 2011 and 2013). The Republic of Albania joined NATO in 2009 and was granted "candidate status" by the European Council in recognition of the efforts Albania enacted with hopes of integrating fully with the EU in 2014-2020 (European Commission, 2014; Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010).

Albania's health care system is primarily public, with the state providing a majority of the health services, including prevention, diagnosis, treatment, and public health initiatives (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). Specialized clinics, as well as pharmaceutical and dental services, make up the growing private health care sector within the country, which are largely located in the capital of Tirana and other major cities (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). In order to determine whether children have access to health services, public health research often looks to vaccination rates as a proxy for evaluating potential disparities in the health system (Gavi, n.d.). In Albania, the 2008-2009 Demographic and Health Survey revealed that 95% of children between 18-29 months received all of the necessary immunizations, with less than 1% receiving no immunizations at all (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). These findings suggest that most young children in Albania have access to primary care services.

Study Aims

The current study examined the internal consistency, validity, and reliability of the M-CHAT-R/F-A in urban centers of Albania. Item-level analyses of the M-CHAT-R/F-A were also conducted to evaluate the ability of individual items to predict ASD symptoms in a low-risk sample of Albanian toddlers. Finally, the study examined potential modifications to the screener

within the current sample, with the goal of proposing a screening tool that best estimates probability of ASD while also minimizing false positive rates, towards the goal of maximizing effectiveness while reducing the time and cost of screening.

Hypotheses

1. The M-CHAT-R/F-A will demonstrate internal consistency and positive predictive values (PPV) comparable to previous studies involving international translations of the M-CHAT and will demonstrate good clinical utility to screen for ASD in Albanian speaking communities in Albania and neighboring regions.

2. M-CHAT-R/F-A items related to joint attention and social referencing will best predict ASD symptoms for children who screen positive on the M-CHAT-R/F-A, but will not provide more predictive information about diagnosis than the total score.

3. The original M-CHAT-R/F-A scoring system of three failed items indicating a positive screen on Stage 1 and two failed items on the Stage 2 will continue to best differentiate between groups.

Methods

Participants

This study examined a sample of toddlers screened with the M-CHAT-R/F-A in urban centers of Albania. The sample included 2,594 toddlers between the ages of 16 and 36 months screened with the M-CHAT-R-A at their pediatrician's office during a WCC visit, with a mean age at screening of 24 months (SD = 2.83 months; see Figure 1). Pediatricians from 18 sites in Albanian speaking communities participated in screening. A total of 3,133 children were screened using the M-CHAT-R-A screening tool during their WCC visits at one of 45

participating pediatrician offices in urban areas of Albania, primarily located in Tirana. However, 539 children were excluded for the following reasons: chronological age (CA) less than 16 months (*n*=211), CA greater than 36 months (*n*=56), no CA obtained due to missing birthdate or date of screening (*n*=257), and 15 children received follow up phone interviews as a matter of clinical concern, but did not meet threshold on the screener and were not evaluated. Toddlers were only included if their parents spoke Albanian and provided informed consent at time of screening.

Of the 2,594 toddlers included in the current sample, 50.1% were male (*n*=1,300), 49.7% were female (*n*=1288), and 0.2% did not have a sex indicated on their screener (*n*=6; see Table 2). Ethnicity data was not requested on the screening tool because Albanian collaborators raised concerns regarding the sensitivity of asking for such information and the impact this might have on participation rates. Ethnicity data for all children evaluated was collected (*n*=20), which indicated that 100% of children evaluated were identified as ethnic Albanian. Maternal education for toddlers was collected at time of screening, indicating that 1% (*n*=32) of mothers received no formal education or did not complete primary school, 23% (*n*=610) did not complete secondary school, 26% (*n*=683) completed secondary school, 0.1% (*n*=3) completed some college, 30% (*n*=781) obtained a college or university degree, and 1% (*n*=26) had advanced degrees. Maternal education was not reported on 18% (*n*=459) of screeners. According to a 2008-2009 survey conducted in Albania, the screened sample represents a higher level of maternal education than in the population at large, as results of this survey indicated majority of women in Albania (49%) have attained 8 years of education, with 1% reporting no education or incomplete primary school, 36% completing secondary school, and 13% with a university education or higher (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro,

2010; see Table 1). Chi square analyses were conducted comparing maternal education from the current sample with that of the general population in Albania. The results indicated that more mothers in our sample endorsed having a secondary education or greater in the current screened sample compared to the general Albanian population (X^2 (1, N = 1,452) = 13.36, p = 0.003).

Procedures

The Institutional Review Board at the University of Connecticut (UConn) provided oversight and approval of this study. No ethical review boards existed in Albania at the time of this study, but Albanian personnel took UConn ethics training and passed the ethics exam. Children were enrolled in the study when screened using the M-CHAT-R/F-A during a WCC visit at their pediatrician's office. Investigators provided pediatrician sites with Albanian informed consent papers, a demographic form, the M-CHAT-R-A screeners, and instructions to screen children between the ages of 16 and 36 months. All documents were translated from English to Albanian, back-translated into English, reconciled by a bilingual and bicultural child psychiatrist (ICR), and approved by the UConn IRB. An Albanian investigator (AC), who is a child psychiatrist, also provided pediatrician training on characteristics of ASD, as well as on how to administer the M-CHAT-R-A in their offices. Parents could decline participation at any point by leaving the M-CHAT-R and demographic forms blank. We are unable to determine attrition based on refusal at this point in the study, because pediatrician sites did not track rates of refusal. If a parent gave consent, they then completed a demographic sheet, which included the child's name, date of birth, sex, and date of screening. It also provided space to report caregiver's name, their relationship to the child, contact information, and the mother's level of education.

Screening sites then sent completed demographic forms and M-CHAT-R/F-As to the study site at the Albanian Regional Center for Autism in Tirana, Albania, affiliated with the Albanian

Children's Foundation, where they were processed and scored. Study staff called parents whose children screened positive on Stage 1 of screening to conduct the Stage 2 follow up interview, during which time both items failed and items left blank were queried. At this point in the study, attrition (*n*=126) was determined when staff was unable to contact parents over the phone or when parents refused to participate in Stage 2 of the screening process. If a child continued to screen positive on the M-CHAT-R/F-A after the interview, meaning the child continued to fail two or more M-CHAT-R/F-A items, the parents were offered a diagnostic evaluation free of charge. Parents were free to decline this evaluation. If severe sensory or motor impairments (e.g., blindness or hearing-impaired) precluded their ability to be evaluated using the study instruments, children were not offered the evaluation.

Diagnostic evaluations were conducted either by a child psychiatrist, a pediatrician, or an experienced therapist from the Regional Center. The children were assessed using the following measures: Autism Diagnostic Observation Schedule (ADOS), Parent Evaluation of Developmental Status (PEDS), PEDS: Developmental Milestones (PEDS: DM), Assessment of Basic Language and Learning Skills-Revised Edition (ABLLS), and a semi-structured parent-report interview assessing presence of ASD symptoms. Clinicians administering the ADOS completed reliability training and received distance supervision from a child psychiatrist with extensive experience in ADOS administration, training in the United Kingdom, and who was native Albanian speaking (ICR). For the current study, outcome was assessed based on ADOS classifications (i.e., ASD or non-ASD). All documents were translated and back-translated, except for the ABLLS, which was filled out by Albanian therapists with good English reading skills.

Parents were then provided verbal feedback and a brief written summary from the testing staff about their child's testing results.

Measures

In the current study, measures directly completed by parents were translated to Albanian with authors' permission and back-translated to English under the supervision of a bilingual Albanian study consultant who was a child psychiatrist practicing medicine in the United Kingdom (ICR). All study authors then approved the back-translatin for each translated measure.

Modified Checklist for Autism in Toddlers, Revised Version with Follow up in Albanian (M-CHA-T-R/F-A; Robins et al., 2014) is an Albanian translation of the M-CHAT-R/F a modified version of the M-CHAT.

Autism Diagnostic Observation Schedule (ADOS). The ADOS is a standardized semi-structured interview using play-based methods to assess for the presence of ASD symptoms (Lord, Risi, Lambrecht, Cook, Leventhal, DiLavore, et al., 2000). The ADOS examines behavior observed during the interview across four domains: Communication, Reciprocal Social Interaction, Play, and Repetitive Behaviors. Scoring algorithms follow this four domain structure, with cut-off scores that assign ASD diagnosis for the Communication and Reciprocal Social Interaction domains (Lord et al., 2000). The ADOS was found to have strong interrater-reliability (mean weighted kappas, $M\kappa_W$) for both Modules 1 and 2 (Lord et al., 2000), with inter-rater agreement for assigning ASD vs. non-spectrum diagnoses at 100% for Modules 1 and 3 and 91% for Module 2 (Lord et al., 2000). The Communication and Social domains have been found to have excellent stability for test-retest reliability and good stability for the Stereotyped Behaviors and Restricted Interests domain (Lord et al., 2000).

Parent Evaluation of Developmental Status (PEDS; Glascoe, 1997). The PEDS is a standardized, 10 item parent-report screening measure that assesses a child's risk for problems in receptive and expressive language abilities, social-emotional functioning, behavior, self-help skills, global/cognitive functioning, current health, and/or fine- and gross-motor skills based on parent concerns reported during pediatrician visits. The first question on this measure is an open-ended question asking parents to report any concerns they might have about their child, and the remaining questions inquire more specifically about parental concerns in each of the developmental areas listed above. A child's level of risk for developmental difficulties is determined based on the number of reported parent concerns for items that have been shown to be predictive of developmental issues, or "predictive concerns," compared to the number of "non-predictive" concerns, which may not require further referrals. Once a child is assigned a risk category based on the number of "predictive concerns" reported, the PEDS provides suggestions on "paths" for follow-up at each level of concern. The PEDS is highly correlated with intellectual functioning, language skills, academic performance, and adaptive behavior abilities (Glascoe, 2003). It has also demonstrated sensitivity estimates between 74-78% and specificity estimates between 70-80% in detecting developmental issues in children between the ages of 0-8 years (Glascoe, 2003). The current study used a translated version of the PEDS with permission from the authors.

PEDS: Developmental Milestones (PEDS: DM) (Glascoe & Robertshaw, 2008). The PEDS: DM "Assessment Level" was used in tandem with the PEDS during diagnostic evaluations in the current study. This measure was designed to provide age-equivalences and percentages of delay across developmental domains in a format that is quickly and easily administered in pediatrician's offices. Percentage of delay greater than 25% indicates clinically significant

developmental concerns, requiring follow-up. The PEDS: DM relies on parent report and/or clinical observation to measure skills in domains consistent with those assessed through the PEDS, which were listed previously. The measure consists of one item per skill domain for each age range for a total of eight items that provide the most predictive information about future developmental trajectory. The PEDS: DM demonstrates excellent sensitivity and specificity overall at 83% and 84%, respectively, with both sensitivity and specificity ranging from 80-93% in children between 16-30 months, the age range of the current study (Brothers, Glascoe, & Robertshaw, 2008). Interrater reliability of the original, English version measured using Guttman's λ coefficient was .98 and test-retest reliability was between .98 and .99 (Brothers et al., 2008). "Readability" has also been assessed using the Flesch–Kincaid index, indicating that the measure is at the 1.8 grade reading level (range 1.1. to 2.6) in English (Brothers et al., 2008). Using the PEDS together with the PEDS: DM provides a more complete assessment of developmental level, as it combines parent concerns with specific developmental milestones (Glascoe & Robertshaw, 2008). The current study used a translated version of the PEDS: DM during evaluations with permission from this measure's author.

Data Analytic Plan

Validation of the M-CHAT-R/F-A was assessed through internal consistency analysis, using Chronbach's alpha, and PPV for autism at each level of screening. Item level analyses were conducted using discriminant function analyses to determine whether particular items on the M-CHAT-R/F-A were predictive of diagnostic outcome (e.g., "ASD" or "Other developmental concerns"). Potential modifications to improve performance of the M-CHAT-R/F-A were then examined.

Results

Screening Results (M-CHAT-R-A)

The current study examined the screening and evaluation results of a sample of 2,594 toddlers between 16 and 36 months of age to determine the clinical utility of the M-CHAT-R/F-A in Albania. Of these children, 253 (9.75%) toddlers screened positive on the M-CHAT-R-A. Follow up interviews were conducted using the M-CHAT-R/F-A with 127 (50.2%) of the screened positive toddlers. The additional 126 were lost to follow-up due to parent refusal or being unable to contact. A total of 26 of the 127 toddlers who received the follow-up interview (21%) continued to screen positive on the follow-up (M-CHAT-R/F-A), qualifying them for a free developmental evaluation. Of those 26 eligible for an evaluation, 20 (77%) attended, and 6 parents of screened positive toddlers refused the evaluation (see Fig 1 for a flow chart of results; see Tables 3, 4, and 5 for a summary of total scores by screening outcome).

A total of 20 children were evaluated, 19 of whom received an ADOS. Of the 19 toddlers who received an ADOS, 17 met criteria for ASD (n=3) or Autism (n=14) and 2 were classified as Non-ASD. (see Figure 1) One child moved out of the country before receiving the ADOS, thus this child's ADOS classification could not be determined. This child's data were removed from further analyses.

Characteristics of Evaluated Sample

Mean age at screening for evaluated children was 26.05 (SD=5.30) months (see Table 6). Of the 20 children evaluated, 75% were male (*n*=15) and 25% female (*n*=5). Maternal education levels indicated that 30% of mothers (*n*=6) had not yet finished secondary school, 35% (*n*=7) completed secondary education, 10% (*n*=2) had a college or university degree, and 25% (*n*=5) did not report this information. Levels of maternal education within this sub-group of the larger

sample were consistent with those reported in the 2008-2009 Albanian survey (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). All of the M-CHAT-R-A screeners (*n*=20) were completed by mothers at their child's pediatrician offices.

The average total score on the M-CHAT-R-A screening tool for toddlers who met ADOS ASD or Autism criteria was 11.71 (SD=4.07), while the average total score for the two Non-ASD children was 9.50 (SD=2.12). On the follow up interview, or M-CHAT-R/F-A, toddlers who met ASD criteria had a mean score of 9.18 (SD=3.59) and toddlers who did not meet ASD criteria had a mean score of 6.00 (SD=2.83; see Table 7 for summary of scores for each evaluated child). Statistical analyses were not run on these groups, because of the small number of children in the non-ASD group (n=2).

Evaluation Data: ASD Group. A total of 17 children met criteria for ASD, with 3 children meeting criteria for Autism Spectrum and 14 meeting criteria for Autism. As expected, scores for children who met Autism criteria were higher (more symptomatic) than children in the Autism Spectrum group on all ADOS domains (see Table 2).

Each evaluated child was assessed using the PEDS and PEDS: DM to estimate developmental abilities. The results from the PEDS, a screening measure assessing risk for developmental problems in children, indicated that the toddlers who met ADOS criteria for Autism had an average of 2.79 (SD=0.58) predictive and 2.50 (SD=1.61) non-predictive concerns. In comparison, the children who met ASD ADOS criteria had an average of 2.33 (SD=0.58) predictive concerns and 2.67 (SD=0.58) non-predictive concerns. Each toddler in the Autism and ASD groups received a "Path A" classification on the PEDS screening tool based on the number of predictive concerns reported by parents, which indicates high risk of developmental disability and need for follow up. On the PEDS: DM, a percentage of delay score

greater than 25% indicates a clinically significant level of delay in a particular area. As a group, the average percentage of delay score for toddlers who met ADOS criteria for Autism exceeded 25% in all but one domain assessed. The average Gross Motor score for the Autism group was just below 25% at 24.25% (SD=20.87%). Toddlers in the ASD group had an average percentage of delay score greater than 25% in the Self-Help, Receptive Language, Expressive Language, and Social Emotional domains. The average percentage of delay values for the Fine and Gross Motor domains were both less than 25% (M=15.52%, SD=17.68 and M=4.81%, SD=32.16%, respectively). The ASD and Autism groups demonstrated significant delays across developmental domains, with the exception of fine and gross motor skills.

Evaluation Data: Non-ASD Classified Group. Two evaluated children were classified as Non-ASD on the ADOS. The mean score for these children on the Communication domain of the ADOS was 3.50 (SD=2.12), an average score that is above the cut-off for ASD [ADOS ASD cutoff score = 2]. In the Reciprocal Social Interaction domain, the mean score was 3.00 (SD=1.41), with a combined Communication and Reciprocal Social Interaction average score of 6.50 (0.71), both average scores falling below the ASD cut-off [ADOS ASD cutoff score = 7]. The Non-ASD classified children had an average score of 3.00 (SD=1.41) in the Play domain and 1.00 (SD=1.41) in the Restricted Interests and Repetitive Behavior (RRB) domain. Non-ASD toddlers had lower scores for each ADOS domain than both the ASD and Autism groups, with greater differences between the ASD/Autism and Non-ASD toddlers in the Reciprocal Social Interaction and Repetitive Behavior domains. Smaller differences between groups were found in the average Communication and Play domains (See Table 7 for summary of scores for evaluated children).

On the PEDS, the average number of predictive concerns for the Non-ASD classified children was 3 (SD=0) with an average of 3.5 (SD=0.71) non-predictive concerns. The Non-ASD toddlers demonstrated, on average, significant percentage of delays in each domain assessed, with the exception of the Gross Motor domain (M=5.02%, SD=12.54%). The 2 Non-ASD toddlers had higher average numbers of predictive and non-predictive concerns than either the ASD or Autism groups. Similar to the ASD and Autism groups, both children in the Non-ASD group also met PEDS criteria for "Path A." On the PEDS: DM, the Non-ASD group had lower overall percentage of delay scores than both the ASD and Autism groups for the Self Help (M=41.35%, SD=23.12%), Receptive Language (M=47.12%, SD=39.43%), and Social Emotional (M=35.04%, SD=10.27%) domains, indicating less severe impairments in these areas. However, the average percentage of delay score for the Expressive Language domain (M=54.49%, SD=17.22%) for these children indicated greater delays when compared to the ASD group and slightly less delayed expressive language abilities compared to the Autism group. In terms of Gross Motor skills, the average percentage of delay score for the two Non-ASD toddlers was slightly higher than the ASD group (M=5.02%, SD=12.54%), though lower than the Autism group. Finally, these two toddlers received greater overall percentages of delay scores in the Fine Motor domain than both the ASD and Autism groups, revealing greater deficits in these areas (M=39.42%, SD=20.47%). These results suggest that while these two children did not meet Autism or ASD criteria on the ADOS, they demonstrated significant developmental delays, leading them to be classified as having non-ASD related developmental delays by evaluating clinicians.

Lost to Follow Up Toddlers. As stated previously, 50% of children who screened positive on the M-CHAT-R-A screener did not receive a follow up interview (see Table 8 for a summary

of demographic characteristic comparisons). Chi square analyses comparing toddlers who received the M-CHAT-R/F-A with those lost to follow up indicated the groups did not differ significantly in terms of sex (X^2 (1, N = 253) = 1.434, p = 0.231) or maternal education (X^2 (5, N = 253) = 10.287, p =0.067). Age at screening for both groups was not normally distributed, with skewness of 0.531 (SE = 0.215) and kurtosis of 0.898 (SE = 0.427). Mean age at screening for toddlers lost to follow up and toddlers who received the M-CHAT-R/F-A was compared using a Mann-Whitney U test, revealing no significant differences in age at screening between those who received a follow up interview and those who did not (U=7889.0, Z = -0.198, p = 0.843). M-CHAT-R-A total scores for the two groups were also non-normally distributed, with skewness of 1.52 (SE = 0.215) and kurtosis of 1.45 (SE = 0.427). A Mann-Whitney U test indicated total scores on the M-CHAT-R/F-A for children who did not receive follow up via phone interview did not differ significantly from those whose parents completed the first stage of screening (U=7641.00, Z = -0.672, p = 0.501). Therefore, attrition seems attributable to other factors, but not the demographic factors assessed or the M-CHAT-R-A original score.

Refused Evaluation. Six children did not receive evaluations due to parent refusal. To determine whether these toddlers differed significantly from the children who received an evaluation, independent-sample t-tests and chi square analyses were used to compare the two groups (see Table 9 for summary of demographic information). Results of these tests indicated that toddlers who did not receive evaluations did not differ significantly on chronological age at screening ($t(24) = -1.067$, p = 0.297), maternal education (X^2 (3, N = 26) = 3.795, p = 0.284), sex (Fishers' exact, p = 0.529), or total score on the M-CHAT-R-A ($t(24) = -0.190$, p = 0.273). Thus, attrition at this stage of the study also seems nonsystematic.

Examination of M-CHAT-R/F-A

Reliability. Internal reliability for the 20 M-CHAT-R-A screener items was good (α = 0.737). Reliability for the M-CHAT-R/F-A was also calculated. Results indicated the follow up interview had excellent reliability (α = 0.935).

Positive Predictive Value (PPV). The current study utilized the original scoring criteria outlined in the Robins et al. (2014) validation of the M-CHAT-R/F in the U.S., with screen positive cut off scores of $\geq$3 on the screening tool and $\geq$2 on follow up. The PPV of the M-CHAT-R-A before follow up for any developmental delay was 0.16. The PPV of the screening tool for ASD or Autism based on ADOS results was 0.14. These figures do not include children who were lost to follow up for the phone interview or refused evaluation. After follow up interview, the PPV of the M-CHAT-R/F-A for ASD or Autism was 0.895 (95% CI: 0.65-0.98) and 1.00 (95% CI: 0.79-1.00) for any developmental delay.

Other Performance Factors. Screened negative rate of the original screener was 90%. Data for determining whether children who screened negative did not meet criteria for ASD were unavailable, due to limited resources for evaluating children who screened negative on Stage 1 of the M-CHAT-R/F-A. The percent of children who screened positive initially but then went on to screen negative on the follow-up was 79.5%, with 101 of 127 screened positive children passing the follow up interview during Stage 2 of the screening process, reinforcing the necessity of reducing false positives with the follow-up interview. The combination of limited data in this area and the small sample size of evaluated children precluded examiners from using ROC curve analyses to evaluate sensitivity and specificity more precisely.

Prevalence of ASD Detected

Although the current study was not intended to be a population-based study of prevalence, we can estimate the percent of the population that screening and diagnostic procedures detected.

Seventeen children met ADOS criteria for ASD out of a total 2,594. However, approximately 50% of the screened positive toddlers (*n*=126) were lost to follow up, with no significant differences found between children who received the M-CHAT-R/F-A and those who were lost to follow up. We therefore assumed these children would have met criteria for Autism at a similar rate if evaluated. Thus, prevalence was estimated by assuming that 34 children would have met criteria for ASD out of the total sample (*n*=2594), resulting in an estimate of 1.31% of the population diagnosed with ASD.

Item Level Data

Item Response Stability. The frequency with which item responses changed between initial screening and follow up on the M-CHAT-R/F-A was also examined. Items that changed the least between the 2 stages of screening included 8, 1, and 6 (see Table 10). On these items, screen positive (failed) responses changed to screen negative (passed) at rates of 42%, 43%, and 45%, respectively. The most dramatic change in initial failure rates on the screener and passing rates after follow up was found for Items 2, 5, and 12, with rates of failure decreasing by 98%, 89%, and 98%, respectively (see Table 10). These three items were the only reverse scored items on the M-CHAT-R-A, suggesting that these items may have been particularly difficult for parents to understand. Additionally, Items 2, 5, and 12 did not involve social interaction or communication. Instead, the items ask about ASD symptoms that may have been worded in a way that made them less easy to understand for parents in Albania (e.g., "Does your child make unusual finger movements near his/her eyes?"; see Table 11 for content of these items).

Missing Items on the M-CHAT-R-A. Parents completed the screener by marking "YES" or "NO" to each question; however, parents also left items blank. The frequency with which items were left blank ranged from 4-192 instances. Item 6, "Does your child point with one finger to

ask for something or to get help?" was left blank on 192 or 7% of all completed screeners, a number 7 times more frequent than the next most frequently blank item. The second most frequent, Item 7, "Does your child try to get you to watch him or her?" was left unanswered on 27 (1%) screeners. Finally, Item 3, "Does your child play pretend or make-believe?" was left blank on 20 (0.8%) screeners. The first item of the M-CHAT-R-A, which asked whether the child follows their parent's point to something across the room, resulted in the fewest blank responses with only 4 (0.15%) left blank of all completed screeners (see Table 11). Of toddlers who received both stages of the screener, the majority of missing items were passed during Stage 2 of the M-CHAT-R/F-A (see Table 11).

Discriminant Function Analysis. A discriminant function analysis (DFA) was used to evaluate the ability of the 20 items on the M-CHAT-R-A to predict ADOS outcome of ASD. The sample was divided into two categories for comparison, with the first group comprised of toddlers who met ASD and Autism criteria on the ADOS (n=17) and the second consisting of toddlers who screened negative on the M-CHAT-R-A or M-CHAT-R/F-A (n=2442). DFA revealed one function, consisting of 12 items with strong predictability for ASD (see Table 12 for items and standardized canonical discriminant function coefficients). These items consisted primarily of questions related to social interaction, including nonverbal communication, response to name, social smiling, shared interest in items or activities, receptive language abilities, eye contact, imitation, and interest in other children (see Table 12). Item 1, "If you point at something across the room, does your child look at it?" was most predictive of meeting ASD or Autism criteria on the ADOS, with Item 10, "Does your child respond when you call his or her name?" being the second most predictive of ASD outcome. Notably, Item 1 was one of the items that changed least frequently between initial screening and follow up. The least predictive

items indicated by the DFA consisted primarily of the three foil items (Items 4, 13, and 20) assessing symptoms unrelated to ASD and the three reverse scored items (Items 2, 5, and 12, see Table 12). Of these 6 items, 5 of the least predictive items identified by the DFA were also items that changed most frequently during Stage 2 of screening, which further suggests that these items may be particularly problematic. Though not a foil or reverse scored item, Item 6, "Does your child point with one finger to ask for something or to get help?" was also one of the least predictive items (0.086), perhaps because of shared variance with Item 1 involving pointing for joint attention. Notably, Item 6 was the item most frequently left blank (*n*=192). However, when administered during Stage 2 of screening, 97% of toddlers passed Item 6, suggesting that this item may not have performed as expected. Item 12, "Does your child get upset by everyday noises?" was revealed to be the least predictive of outcome (-0.035). This item is a reverse scored item that was rarely left blank on the initial stage of screening (*n*=15 or 0.6%). While 47% of all parents endorsed this symptom on the screener, 98% of the toddlers who received Stage 2 of the M-CHAT-R/F-A went on to pass the item when administered over the phone and parents were asked for additional information (see Table 10). Based on its instability over the two stages of screening and minimal predictive value, Item 12 was the poorest performing item on the M-CHAT-R/F-A.

Removal of Three Problematic Items

The performance of the M-CHAT-R/F-A was then reevaluated after omitting the three items most frequently changed over the two stages of screening and found to be the least predictive by DFA (Items 2, 5, 12, see Figure 2 for Revised Flow Chart). Reliability for Stage 1 of screening increased from 0.737 on the 20-item version to 0.886 without the three items, and remained stable for Stage 2 at 0.931 without the items, with the full version of the screener being 0.935.

Results also indicated that screened negative cases increased to 96% on Stage 1 of screening compared to 90% of the original M-CHAT-R-A cases (see Table 13), with 146 fewer children screening positive (*n*=107) without the three items. Of the 107 screened positive toddlers, 40% continued to screen positive when Items 2, 5, and 12 were removed, compared to 20% of toddlers on the M-CHAT-R/F-A. With higher reliability and fewer false positives at Stage 1, as well as a higher percentage of children who continued to screen positive after Stage 2, the revised version of the M-CHAT-R/F-A may be a more effective screening tool for ASD in the current sample. Of course, it is not possible to state this with confidence until the revised screener is tested with these changes already made.

In terms of evaluations under this proposed version of the screener, 24 children were eligible for an evaluation, 20 of whom received an evaluation under the original version and 4 refused (see Table 13 for comparison with original findings). Only 2 cases determined to require an evaluation by the original M-CHAT-R-A no longer met criteria when the three items were removed, both of whom refused the evaluation after follow up, and whose diagnosis is therefore unknown. Roughly 45% of the 107 children who screened positive without the three items included in the screener were lost to follow up, a figure roughly consistent with the percentage lost under the original version used in this study.

Results also indicated a possible alternative cut-off score for Stage 1 of M-CHAT-R/F-A without the three problematic items. Examination of the frequency of total scores by screening outcome revealed that no children who received an evaluation and subsequent diagnosis scored lower than 5 items failed on Stage 1 of the currently suggested version of the screener. This suggests a possible alternative scoring criteria of 5 total items failed on Stage 1 of the screener when Items 2, 5, and 12 are omitted, with the continued best practice scoring criteria of 2 failed

items on Stage 2. However, further research is needed to evaluate the performance of this version of the screener, with the proposed items removed and 5-point cut-off score.

Discussion

M-CHAT-R/F-A Performance

This is the first study conducted outside of the U.S. to examine the utility of a translated version of the recently validated M-CHAT-R/F. Both stages of the M-CHAT-R/F-A were translated and adapted for use in Albanian-speaking communities to detect ASD in toddlers aged 16-36 months at their WCC visits. Results indicted excellent reliability for the 2-Stage M-CHAT-R/F-A ($\alpha = 0.935$) in the current sample, a figure higher than that of the original M-CHAT-R/F ($\alpha = 0.79$; Robins et al., 2014).

Of the 2,594 children initially screened with the M-CHAT-R-A, 20.5% continued to screen positive during Stage 2 of screening with the M-CHAT-R/F-A. PPV for ASD at Stage 1 of screening was 0.14. This figure increased dramatically after Stage 2 follow-up was conducted, yielding PPVs of 0.89 for ASD or Autism and 1.00 for any developmental delay or concern. Comparatively, PPV for any developmental concern on the M-CHAT-R/F in the U.S. was 0.946 (Robins et al., 2014). PPV for ASD in the current sample was substantially higher than the original M-CHAT-R/F study (0.475). However, the current sample (n=2594) was smaller than that of the original U.S. study (n=15612); thus the current results may be an overestimate of the true positives existing within the Albanian speaking population in Albania (n=2549 in Albania vs. n=15612 in U.S.; Robins et al., 2014). Prevalence of ASD in Albania was estimated at 1.31%. Current estimates in the U.S. propose a similar rate of 1.47%, while global estimates are

much lower, suggesting that 0.62% of the world's population has ASD (Center For Disease Control, 2015; Elsabbagh et al., 2012).

While the current study followed several of the original M-CHAT-R/F validation study procedures, several key procedural modifications were made. In the U.S., toddlers who screened positive on the M-CHAT-R with a score of 7 or higher were automatically provided an evaluation and those with scores between 3 and 6 were given a follow up interview. The current study, however, followed the original authors' suggested total score criteria of three missed items on Stage 1 of screening in order to ensure that all screened positive children received follow up interviews, allowing investigators as much data possible to evaluate the 2-Stage screening process in Albania. The PPV for any developmental concern in high-risk children (those who scored at or above a 7 on the screener) in the U.S. was 100%, which is consistent with our findings for the entire current sample. It may be that the sample in this study included toddlers at higher risk for ASD, with parents of more delayed children being more likely to follow up with professionals offering assistance, which supports this comparable PPV. However, comparisons made in the current study between children lost to follow up or those whose families refused evaluation also revealed that the current sample was not confined to high-risk, high scoring toddlers, as average total scores did not significantly differ between these groups and those who completed the study. Instead, the comparability between the Albanian and high-risk U.S. sample may be partially attributable to procedural differences, in that the PPV of ASD for the M-CHAT-R/F included only "medium-risk" toddlers with total Stage 1 screening scores between three and six, while the current sample's PPV results included all screen positive children, for whom the average M-CHAT-R-A score was 11.25 (SD=4.83).

The percentage of children requiring follow up after initial screening in Albania was 9.75%, which is slightly higher than the 6% screen positive rate in the U.S. in the medium-risk category (Robins et al., 2014). These findings may again be due, in part, to differences in study procedures, with high scoring U.S. toddlers not being included in the 6%. Of those who screened positive, 26 toddlers (20.5%) continued to fail Stage 2 of screening, qualifying them for a free developmental evaluation. In contrast, 79.5% of toddlers who screened positive at Stage 1 of the M-CHAT-R/F-A no longer screened positive after Stage 2 follow up, a large figure given the resources required to conduct follow up interviews with caregivers (e.g., money needed to call caregivers, time required by staff to administer items via phone and solicit examples, making multiple calls to families before contact made, etc.). The number of false positives on initial screenings in the total low-risk sample in the U.S. was 63% (Robins et al., 2014). While a high percentage, this figure indicates fewer false positives at Stage 1 of screening in the U.S. than in the current sample.

The necessity of the follow up interview for increasing sensitivity and specificity is well demonstrated in the M-CHAT and M-CHAT-R/F literature (Robins et al., 2001; Kleinman et al., 2007; Chlebowski et al., 2013; Robins et al., 2014). Unfortunately, the resources needed to follow up with families post screening may impose a strain on health care providers who screen for ASD during WCCs, particularly those in resource-limited settings. Attrition rates such as those found in the current study (50% of screened positive toddlers lost to follow up and 23% refused evaluation after continuing to screen positive on follow up interview) further contribute to the difficulties posed by the 2-Stage process. Attrition due to caregiver inaccessibility (e.g., phone number changed, moved, etc.) can preclude at-risk children from receiving further services. Moreover, cultural differences in expectations regarding child development and in

attitudes surrounding stigma of mental illness may also contribute to parent refusal to discuss concerns over the phone or to receive a developmental evaluation for their child (Koh et al., 2014; Samadi and McConkey, 2015). The problem of attrition across the 2-Stage screening process is not confined to the current sample. In the U.S., 18% of children were lost to follow up after Stage 1 of screening and 37% either refused or did not complete evaluations following Stage 2 (Robins et al., 2014). Studies examining the 2-Stage screening process with translated and/or culturally adapted versions of the M-CHAT screener in Japan and South Korea also reported high rates of toddlers lost to follow up between initial screening and the follow up interview, with both studies reporting they were unable to follow up with 39% of their screened samples (Kamio, Inada, Koyama, Inokuchi, Tsuchiya, & Kuroda, 2014; Seung, et al., 2015). Kamio and colleagues (2014) proposed that the high false positive rate found in their study (17%) may have been a contributing factor to attrition in their sample. The authors suggested that Japanese caregivers may have experienced "unnecessary" distress as a result of their child being identified incorrectly as being at risk for ASD, which may have impacted participation rates in follow up (Kamio et al., 2015). The current study did not find any significant differences between toddlers who were lost to follow up or refused evaluations and the toddlers who received Stage 2 follow up or evaluations based on total scores on either Stages 1 or 2 on the M-CHAT-R/F-A, average age of toddler, sex of child, or maternal education.

Researchers continue to examine alternatives to established screening protocols that will best suit the needs of each population and context, including measures for decreasing attrition. Studies such as one investigating the M-CHAT's performance in Turkey addressed this issue, as well as a cultural difference found in caregiver response patterns there, by offering follow up interviews in-person at the time of screening (Kara et al., 2012). Currently, the original authors

of the M-CHAT and M-CHAT-R/F are examining the utility of an electronic version of the M-CHAT-R screener, with follow up being administered immediately following the screener electronically. If found to be a valid administration process of the M-CHAT-R/F in the U.S., dissemination and evaluation of this technology should be examined in other regions, particularly those where higher attrition rates are found. Additionally, modifications to the screening tool itself, as well as to cut off scores, have been shown to improve M-CHAT performance when adapted for use outside of the U.S. setting (Koh et al., 2013; Inada et al., 2011; Wong, Hui, Lee, Leung, Ho, Lau,... & Chung, 2004; Kamio et al., 2015). Such modifications were also explored in the current study and will be discussed in more detail later in the discussion.

Evaluated Sample

Although more formal validity analyses were not conducted due to small group sizes, results indicated that children in the ASD and Autism groups had higher average total scores for both stages of screening with the M-CHAT-R/F-A when compared to the non-ASD toddlers. This suggests there is a relationship between total scores on both Stage 1 and 2 of the screener and final outcome. The non-ASD toddlers received lower ADOS scores than children in both ASD and Autism groups for each ADOS domain, with the greatest differences in scores found between groups for the Reciprocal Social Interaction and Repetitive Behavior domains. These results are not surprising given that these domains represent hallmark features of ASD, and suggest that the non-ASD toddlers demonstrated fewer ASD symptoms than their peers who were classified as having an ASD or Autism.

Initial screening for developmental concerns using the PEDS revealed that parents of each evaluated child reported high numbers of predictive concerns, indicating referral to a

professional for further evaluation was needed for both ASD and non-ASD toddlers. Estimation of developmental abilities using the PEDS: DM revealed significant percentages of delay for most domains assessed in both the ASD and Autism groups, with the exception of the Gross Motor domain. These findings suggest that rather than being delayed more broadly across developmental skill areas, the ASD and Autism groups, on average, demonstrated patterns of delays consistent with their diagnoses (e.g., social/emotional functioning, communication, adaptive). This is consistent with other literature indicating motor skills can be an area of strength for children with ASD (Ventola, Kleinman, Pandey, Wilson, Esser, Boorstein, ... & Green, 2007). These results also suggest that toddlers in the ASD and Autism groups did not screen positive on the M-CHAT-R/F-A as a result of more global delays, but instead due to their constellation of symptoms. The non-ASD children demonstrated more delayed areas of development than the ASD group, though their scores were often lower than the Autism group. Toddlers who did not meet criteria for ASD were found to have greater delays in fine motor skills than both the ASD and Autism groups.

Robins and colleagues (2014) reported that 94.5% of their evaluated sample in the U.S. had significant developmental delays or concerns requiring referral to early intervention specialists. The study authors reported higher average scores on the M-CHAT-R, MCHAT-R/F, and ADOS for the toddlers with ASD compared to those who did not meet diagnostic criteria, a pattern consistent with the findings from the current study (Robins et al., 2014). Mean scores on developmental/adaptive measures were less consistent with the pattern observed in the current sample, with non-ASD toddler groups having higher scores on measures of communication, daily living, socialization, and motor skills than the ASD groups in the original M-CHAT-R/F validation study (Robins et al., 2014).

Results of the DFA supported the initial hypotheses that joint attention items on the M-CHAT-R-A would predict ASD classification (see Table 12). The item that was most predictive of ASD or Autism classification on the ADOS was Item 1, "If you point at something across the room, does your child look at it?" Response to name, social smiling, showing items, and following instructions were each more predictive of outcome than the item pertaining to toddlers seeking caregivers' emotional responses before selecting their own response. Alternatively, the least predictive items for ASD classification in the current sample included those related to pretend play, making unusual finger movements, and oversensitivity to noise. Each of the reverse scored items and foil items were also found to be least predictive of outcome. Finally, contrary to initial hypotheses, findings revealed that Item 6, "Does your child point with one finger to ask for something or to get help?" also poorly predicted ASD classification in the current sample. However, this result may also be related to two other factors noted in the results. First, Item 6 was left blank on 7% of all completed screeners, thus problems in wording may have influenced parents' tendency to not respond. Further, the variance shared between this item and the most predictive item, Item 1, which also inquires about the toddler pointing to items may have also reduced the predictive value of this question, a result also demonstrated by the DFA conducted for the original M-CHAT study (Robins et al., 2001).

These results were further supported by the patterns of responses found on the screener, as well as the stability of responses across the 2-Stage screening process. As might be expected, items frequently left blank on the initial screener or changed consistently during the follow up interview (e.g., from failure to passing) were also the least predictive of ASD outcome. Three items emerged from the findings as particularly problematic (Items 12, 5, and 2, respectively). The poor performance of these items may be partially related to the content of the questions, as

none of the three items inquire about hallmark ASD symptoms. Therefore, it may be that these items were less predictive due to the skills they assessed. In addition, each of these items is reverse scored (e.g., "yes" response indicates ASD symptom presence), thus potential difficulties with wording or the nature of the questions may also have interfered with item performance.

In efforts to better understand these findings, a third-party, native Albanian speaker who was not affiliated with this study was asked to review these three items to suggest potential problems with wording or concept. This consultant suggested that the Albanian translations for "deaf" and "unusual finger movements," from Items 5 and 2, may have been less familiar to some Albanians. Wording selection for Item 12 was not thought to be problematic. While word-choice may be partially responsible for the poor performance of Items 5 and 2, other factors appear to be interfering with the effectiveness of Item 12 as a predictor of ASD in this sample.

In validating the M-CHAT-R/F for use in the U.S., Robins and colleagues (2014) found that total score best predicted ASD outcome, rather than a subset of individual items. In studies investigating translated and adapted versions of the original M-CHAT, several investigators found results more consistent with the current study, suggesting that a number of individual items provided important information about ASD outcome, while total score less predictive of diagnostic outcomes. Although wording was changed for all but three items (Items, 6, 7, and 17) when developing the M-CHAT-R/F, significant overlap in symptoms targeted by the two versions remained. Thus, comparison of current findings with those examining translated versions of the original screener can provide important information for contextualizing M-CHAT-R/F-A results.

Several internationally conducted M-CHAT studies supported current findings regarding predictive value of items on the initial screener. Items 1, 7, 8, 9, 10, 14, 16, 17, and 18 were

found to perform well by multiple investigators across different region of the world, including Japan, China, Turkey, South Korea, and several Middle Eastern countries (Inada et al., 2011; Wong et al., 2004; Kara et al., 2012; Seung et al., 2015; Seif Edin et al., 2008). Each of these items assesses a child's capacity for joint attention and responsiveness to social stimuli, which have also been shown to best differentiate between children with ASD and other developmental concerns in previous literature (Ventola et al., 2007).

International studies examining the performance of the M-CHAT also reported findings related to problematic or less predictive items that were consistent with the current study. Three such studies found Items 12 and 5, the poorest performing items based on current results, to be less predictive in their samples (Canal-Bedia et al., 2011; Kara et al., 2012; Seung et al., 2015). Authors of a validation study for the M-CHAT conducted in Spain proposed that higher failure rates on Items 12 and 5 (Items 11 and 18 on M-CHAT) were possibly due to parents misunderstanding the wording, thereby interfering with their accurate comprehension of the items (Canal-Bedia et al., 2011). In South Korea, Seung and colleagues (2015) reported that parents' responses on M-CHAT items equivalent to M-CHAT-R/F Items 12 and 5 changed more frequently after follow up than others, albeit the change rates reported were lower than the current study. Seung and colleagues (2015) reported rates of change between Stages 1 and 2 of screening at 65.7% and 5.5%, respectively for these two items, while the current study revealed both items changed at a rate of 98%. These authors reported that the higher initial failure rates were potentially a result of misunderstanding item meaning due to "culturally-based" differences revealed during follow up interviews (Seung et al., 2015). In the case of Item 5 involving "unusual finger movements," the authors suggest that South Korean parents may have had difficulty understanding the intent of the question due a common form of dyadic play involving

pointing, (Seung et al, 2015). Further, the South Korean study authors proposed a shorter version of the Korean M-CHAT excluding these two items, as well as the other two reverse-coded items found in the original M-CHAT (the M-CHAT-R/F has three such items), due to the high failure rates in their sample (Seung et al., 2015). A study conducted in Japan similarly proposed creating a short form of the Japanese M-CHAT, also excluding the four reverse scored items on the original screener, suggesting fewer false positives would result (Inada et al., 2011).

Differences were also found between results of the current study and those reported in the literature. Items found to be predictive in Albanian speaking communities, such as social referencing and imitation were found to be similarly predictive in studies conducted in Japan and China; however, they were less predictive in studies conducted in Spain, Turkey, or several Arabic speaking countries (Kara et al., 2012; Inada et al., 2011; Wong et al., 2004; Seif Edin et al., 2008). The converse was also true, with Items 3 (pretend play) and 11 (pointing to request) found to perform poorly in the current Albanian sample, while studies conducted in Japan found both items to be predictive and others in China and Turkey found Item 11 to perform well in their samples (Kara et al., 2012; Inada et al., 2011; Wong et al., 2004).

It is difficult to determine the precise causes for such differences in results between M-CHAT studies conducted across the world. However, as the field of ASD research continues to expand to include other regions of the globe, researchers have begun to explore potential differences in cultural perceptions regarding child development, completion of written screening tools, and of an ASD diagnosis itself. Researchers in this field have found that the salience of items examining social and communication skills differed between countries and cultures (Daley et al., 2013; Perera et al., 2009; Samadi and McConkey, 2015; Kara et al., 2012; Koh et al., 2014; Albores-Gallo, Roldán-Ceballos, Villarreal-Valdes, Betanzos-Cruz, Santos-Sánchez,

Martínez-Jaime, ... & Hilton, 2012). Cultural attitudes toward the propriety of frequent eye contact or pointing between children and adults are suggested as one such example, as parents from these communities may not perceive their children's behavior as atypical if it the behaviors are not encouraged in young children (Albores-Gallio et al., 2012). Other investigators suggest that mothers in LMICs may not notice or interpret their child's delays in communication or social skills as impairments (Daley et al., 2013; Ertem, Atay, Dogan, Bayhan, Bingoler, Gok, ... & Isikli, 2007). The optimal format for collecting information about a child's development may also be culturally dependent. For example, caregiver familiarity with completing paper and pencil questionnaires or with responding to yes/no questions has been shown to influence screening results (Kara et al., 2012; Wong et al., 2004; Inada et al., 2011). Once these factors are better understood in a particular context, modifications to the screener should then be made to improve its utility in that population. More fundamentally, however, the demonstrated performance of the M-CHAT-R and M-CHAT screening tools, which inquire about specific behaviors rather than parent concern, suggests such screeners might yield reasonable estimates of ASD prevalence even with differences between cultural contexts.

Proposed Modifications to M-CHAT-R/F-A

The current study revaluated the M-CHAT-R/F-A, omitting the three most problematic items, Items 12, 5, and 2. These items were selected as poor performing based on their limited predictive value based on DFA results and high frequency of change after Stage 2 follow up. While further research must be conducted to evaluate this shortened version of the screener as a stand-alone measure, analyses of data from the current sample indicate the omission of these items would improve performance by increasing the internal consistency of the screening tool

and decreasing initial screen positive rates, thereby reducing the number of follow up interviews required. Also based on this data, a total cut off score of 5 or greater was suggested.

Limitations

Limitations of the current study include limited ability to determine the number of true negative cases, or the number of children who screened negative but have ASD, due to limited resources being available for evaluating screened negative children. As a result, sensitivity and specificity could not be established. A further limitation of the study relates to the lack of pretesting M-CHAT-R-A items with caregivers to screen for readability and comprehension by Albanian speaking persons without formal experience in child development. Finally, due to the current sociopolitical context in Albania, collecting data on race and ethnicity was discouraged. The current investigators therefore cannot be certain that the population screened in the current study was representative of the Albanian speaking population's racial and ethnic make up. However, given information collected related to socioeconomic status as indexed by maternal education and literature suggesting that more than 99% of children attend WCC in Albania, the current results are considered representative of the Albanian speaking community in Albania (Institute of Statistics, Institute of Public Health [Albania] and ICF Macro, 2010). Results also indicated that the current screened sample reported higher maternal education levels than the general population in Albania. However, this may be due to parents with higher education being more willing to complete screeners. These parents also live in urban settings, such as Tirana, which may also contribute to this higher education level in our sample.

In addition, although a large number of children were screened, only 19 were fully evaluated; it is difficult to collect a larger sample of screened children in a reasonable time frame, but this

makes the conclusions about screening efficacy in need of replication. Finally, the suggested modifications of the screener need to be tested as a stand-alone measure.

Conclusions

The current study revealed strong performance for the use of the M-CHAT-R, Albanian version to detect ASD in toddlers at their WCC visits in Albanian speaking communities. Although there are no hard data on this point, our Albanian medical colleagues report that most children outside of the study are identified as possibly having autism much later than the current screening age (or not at all). Children in the current study received timely referrals for treatment at the Regional Center for Autism. Results of this study support the conclusion that while ASD screening tools may require cultural adaptations to best meet the needs of different contexts, the core symptoms of ASD (e.g., following a point, bringing to show, shared affect, response to name, gaze following) appear to be consistent across cultures.

References

Albores-Gallo, L., Roldán-Ceballos, O., Villarreal-Valdes, G., Betanzos-Cruz, B. X., Santos-Sánchez, C., Martínez-Jaime, M. M., ... & Hilton, C. L. (2012). M-CHAT Mexican version validity and reliability and some cultural considerations. *ISRN neurology*, *2012*.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. *Arlington: American Psychiatric Publishing*.

Anderson, D.K., Liang, J.W., and Lord, C. (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. Journal of Child Psychology and Psychiatry, 55: 485-494. doi: 10.1111/jcpp.12178

Ashwood, K. L., Buitelaar, J., Murphy, D., Spooren, W., & Charman, T. (2014). European clinical network: autism spectrum disorder assessments and patient characterization. *European child & adolescent psychiatry*, 1-11.

Autism Speaks Global Autism Public Health Initiative. (n.d.). Retrieved 5/1/2015 from: https://www.autismspeaks.org/science/research-initiatives/global-autism-public-health#gaph_goals

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry*, *161*(6), 839–843.

Barton, M. L., Dumont-Mathieu, T., & Fein, D. (2011). Screening Young Children for Autism Spectrum Disorders in Primary Practice. *Journal of Autism and Developmental Disorders.*, *42*(6), 1165–1174.

Brothers, K. B., Glascoe, F. P., & Robertshaw, N. S. (2008). PEDS: Developmental Milestones--An Accurate Brief Tool for Surveillance and Screening. *Clinical Pediatrics*, *47*(3), 271–279.

Canal-Bedia, R., García-Primo, P., Martín-Cilleros, M. V., Santos-Borbujo, J., Guisuraga-

Fernández, Z., Herráez-García, L., et al. (2010). Modified Checklist for Autism in Toddlers: Cross-Cultural Adaptation and Validation in Spain. *Journal of Autism and Developmental Disorders.*, *41*(10), 1342–1351.

Center for Disease Control (2014). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010.

Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*,*131*(4), e1121-e1127.

Daley, T. C., Singhal, N., & Krishnamurthy, V. (2013). Ethical considerations in conducting research on autism spectrum disorders in low and middle income countries. *Journal of autism and developmental disorders*, *43*(9), 2002-2014.

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., et al. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Research*, *5*(3), 160–179.

Ertem, I. O., Atay, G., Dogan, D. G., Bayhan, A., Bingoler, B. E., Gok, C. G., ... & Isikli, S. (2007). Mothers' knowledge of young child development in a developing country. *Child: care, health and development*, *33*(6), 728-737.

European Commission. (2014). Albania Progress Report. Retreived from: http://ec.europa.eu/enlargement/pdf/key_documents/2014/20141008-albania-progress-report_en.pdf

European Forum for Democracy and Solidarity. (2014). Albania. Retreived from: http://www.europeanforum.net/country/albania

Gavi. (n.d.). Health equity. Retrieved from: http://www.gavi.org/About/Value/Health-

equity/#sthash.EaG1n7oc.dpuf

Glascoe, F. P. (1997). Parents' evaluations of developmental status: A method for detecting and addressing developmental and behavioral problems in children. *Nashville, TN: Ellsworth & Vandermeer Press Ltd.*

Glascoe, F. P. (2003). Parents" Evaluation of Developmental Status: How Well Do Parents" Concerns Identify Children With Behavioral and Emotional Problems? *Clinical Pediatrics*, *42*(2), 133–138.

Glascoe, F. P. & Robertshaw, N. S. (2008). Parents' Evaluation of Developmental Status: Developmental Milestones (PEDS:DM). Nashville, Tennessee: Ellsworth & Vandermeer Press, LLC.

Herlihy, L. E. (2014). Racial/ethnic and Socioeconomic Differences in Screening Toddlers for Autism Spectrum Disorders Using the M-CHAT.

Institute of Statistics, Institute of Public Health [Albania] and ICF Macro. (2010). Albania Demographic and Health Survey 2008-09. Tirana, Albania: Institute of Statistics, Institute of Public Health and ICF Macro.

Inada, N., Tomonori Koyama, P., Inokuchi, E., Kuroda, M., & Kamio, Y. (2010). Research in Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *5*(1), 330–336.

Johnson CP, Myers SM; American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007;120(5):1183–1215

Kamio, Y., Haraguchi, H., Stickley, A., Ogino, K., Ishitobi, M., & Takahashi, H. (2015). Brief Report: Best Discriminators for Identifying Children with Autism Spectrum Disorder at an 18-Month Health Check-Up in Japan. *Journal of autism and developmental disorders*, 1-7.

Kamio, Y., Inada, N., Koyama, T., Inokuchi, E., Tsuchiya, K., & Kuroda, M. (2014). Effectiveness of using the modified checklist for autism in toddlers in two-stage screening of autism spectrum disorder at the 18-month health check-up in Japan. *Journal of autism and developmental disorders*, *44*(1), 194-203.

Kara, B., Mukaddes, N. M., Altinkaya, I., Guntepe, D., Gokcay, G., & Ozmen, M. (2014). Using the Modified Checklist for Autism in Toddlers in a well-child clinic in Turkey: Adapting the screening method based on culture and setting. *Autism, 18*(3), 331–338.

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*(2), 152-159.

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., et al. (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet*, *378*(9801), 1515–1525.

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., ... & Song, D. H. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry*, *168*(9), 904-912.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., ... & Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of autism and developmental disorders*, *38*(4), 606-615.

Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. *Journal of Autism and Developmental Disorders., 30*(2).

Koh, H. C., Lim, S. H., Chan, G. J., Lin, M. B., Lim, H. H., Choo, S. H. T., & Magiati, I. (2013). The Clinical Utility of the Modified Checklist for Autism in Toddlers with High Risk 18–48 Month Old Children in Singapore. *Journal of Autism and Developmental Disorders.*, *44*(2), 405–416.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders.*, *30*(3).

Mandell, D. S., Listerud, J., Levy, S. E., & Pinto-Martin, J. A. (2002). Race differences in the age at diagnosis among Medicaid-eligible children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*,*41*(12), 1447-1453.

Matson, J. L., Worley, J. A., Fodstad, J. C., Chung, K. M., Suh, D., Jhin, H. K., ... & Furniss, F. (2011). A multinational study examining the cross cultural differences in reported symptoms of autism spectrum disorders: Israel, South Korea, the United Kingdom, and the United States of America. *Research in Autism Spectrum Disorders*, *5*(4), 1598-1604.

National Research Council (US) Committee on Educational Interventions for Children with Autism (2001) Educating Children with Autism. Washington, DC: National Academies Press.

Patel, V., Kieling, C., Maulik, P. K., & Divan, G. (2013). Improving access to care for children with mental disorders: a global perspective. *Archives of disease in childhood*, *98*(5), 323-327.

Patel, V., Chowdhary, N., Rahman, A., & Verdeli, H. (2011). Improving access to psychological treatments: Lessons from developing countries. *Behaviour Research and Therapy*, *49*(9),

523–528.

Patel, V., Flisher, A. J., Nikapota, A., & Malhotra, S. (2008). Promoting child and adolescent mental health in low and middle income countries. *Journal of Child Psychology and Psychiatry*, *49*(3), 313–334.

Perera, H., Wijewardena, K., & Aluthwelage, R. (2009). Screening of 18-24-Month-Old Children for Autism in a Semi-Urban Community in Sri Lanka. *Journal of Tropical Pediatrics*, *55*(6), 402–405.

Posada, Primo, Ferrari, & Martín-arribas. (2007). Research Institute for Rare Diseases, Instituto de Salud Carlos III, Madrid. Retrieved from: https://www.researchgate.net/publication/253814588_European_Autism_Information_System_EAIS_Report_on_the_%27Autism_Spectrum_Disorders_Prevalence_Data_and_Accessibility_to_Services%27_Questionnaire_Q-_EAIS

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). 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., & Dumont-Mathieu, T.M. (2006). Early screening for autism spectrum disorders: update on the modified checklist for autism in toddlers and other measures. *Journal of Developmental & Behavioral Pediatrics*, *27*(2), S111-S119.

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), 131–143.

Rogers, S. J. (1996). Brief report: Early intervention in autism. *Journal of autism and developmental disorders*, *26*(2), 243-246.

Samadi, S. A., & McConkey, R. (2015). Screening for Autism in Iranian Preschoolers: Contrasting M-CHAT and a Scale Developed in Iran. *Journal of autism and developmental disorders*, 1-9.

Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: scarcity, inequity, and inefficiency. *The Lancet*, *370*(9590), 878–889.

Seif Eldin, A., Habib, D., Noufal, A., Farrag, S., Bazaid, K., Al-Sharbati, M., et al. (2008). Use of M-CHAT for a multinational screening of young children with autism in the Arab countries. *International Review of Psychiatry*, *20*(3), 281–289.

Seung, H., Ji, J., Kim, S. J., Sung, I., Youn, Y. A., Hong, G., ... & Youm, H. K. (2015). Examination of the Korean Modified Checklist of Autism in Toddlers: Item Response Theory. *Journal of autism and developmental disorders*, 1-14.

Simmons and Slocum. (2014). Indo-European Languages, Balkan Group: Albania. Retreived from: http://www.utexas.edu/cola/centers/lrc/general/ie-lg/Albanian.html

Soto, S., Linas, K., Jacobstein, D., Biel, M., Migdal, T., & Anthony, B. J. (2014). A review of cultural adaptations of screening tools for autism spectrum disorders. *Autism*, 1362361314541012.

Szatmari P, Bryson SE, Boyle MH, et al. (2003) Predictors of outcome among high functioning children with autism and Asperger syndrome. Journal of Child Psychology and Psychiatry and Allied Disciplines 44(4): 520–528.

UNICEF. (2013). Albania, statistics. Retrieved from: http://www.unicef.org/infobycountry/albania_statistics.html

USAID. (2011). Albania: 2011-2015, Country Development Cooperation Strategy. Retrieved from: https://www.usaid.gov/sites/default/files/documents/1863/CDCS_USAID_Albania.pdf

Ventola, P., Kleinman, J., Pandey, J., Wilson, L., Esser, E., Boorstein, H., ... & Green, J. (2007). Differentiating between autism spectrum disorders and other developmental disabilities in children who failed a screening instrument for ASD. *Journal of autism and developmental disorders*, *37*(3), 425-436.

Wallace, S., Fein, D., Rosanoff, M., Dawson, G., Hossain, S., Brennan, L., et al. (2012). A Global Public Health Strategy for Autism Spectrum Disorders. *Autism Research*, *5*(3), 211–217.

Williams, J. G., Higgins, J. P., & Brayne, C. E. (2006). Systematic review of prevalence studies of autism spectrum disorders. *Archives of disease in childhood*, *91*(1), 8-15.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., ... & Burstein, R. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet*, *382* (9904), 1575-1586.

The World Bank (2012). Newest country classifications. Retrieved from: http://data.worldbank.org/news/newest-country-classifications-released.

The World Bank (2013). Newest country classifications. Retrieved from: http://data.worldbank.org/news/new-country-classifications

The World Bank (2015). How We Classify Countries. Retrieved from: http://data.worldbank.org/about/country-classifications [accessed 15 May 2014]

WHO (2010). mhGAP intervention guide for mental, neurological and substance use disorders in non-specialized health settings: mental health Gap Action Programme (mhGAP).

WHO (n.d.). Metrics: Disability-Adjusted Life Year (DALY). Retrieved from: http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/

WHO (2013). Meeting report: autism spectrum disorders and other developmental disorders: from raising awareness to building capacity. Retrieved from: http://apps.who.int/iris/bitstream/10665/103312/1/9789241506618_eng.pdf

WHO (2015). Albania. Retrieved from: http://www.who.int/countries/alb/en/

Wong, V. C., & Hui, S. L. (2008). Brief report: Emerging services for children with autism spectrum disorders in Hong Kong (1960–2004). *Journal of autism and developmental disorders*, *38*(2), 383-389.

Wong, V., Hui, L. H. S., Lee, W. C., Leung, L. S. J., Ho, P. K. P., Lau, W. L. C., ... & Chung, B. (2004). A modified screening tool for autism (Checklist for Autism in Toddlers [CHAT-23]) for Chinese children. *Pediatrics*, *114*(2), e166-e176.

Zweigenbaum, L., Bryson, S., Brian, J., Smith, I., Roberts, W., Szatmari, P., Roncadin, C., Garon, N. & Vaillancourt, T. (2015). Stability of diagnostic assessment for autism spectrum disorder between 18 and 36 months in a high-risk cohort, Autism Research, DOI: 10.1002/aur.1585

Zweigenbaum, L., Bryson, S., Lord, C., Rogers, S., Carter, A., Carver, L., ... & Yirmiya, N. (2009). Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants. *Pediatrics*, *123*(5), 1383-1391.

Appendix A

Tables

Table 1. World Bank Economy Classifications based on GNI per Capita

Classification	GNI per Capita
Low-Income	$1,045
Lower Middle-Income	$2,012
Upper Middle-Income	$7,901
High-Income	$12,736

Table 2. Sample Demographic Characteristics

Variable	Total
Total Sample	2594
Age at screening, mean (SD), mo	24.16 (2.83)
Sex	
Female	1300
Male	1288
Not reported	6
Maternal Education	
No formal education/incomplete primary school	32
Unfinished secondary school	610
Secondary education	683
Some college	3
College/University Degree	781
Advanced Degree	26
Unknown	459
Person who Completed M-CHAT-R-A	
Mother	2290
Father	211
Other Family Member	42
Unknown	51

Table 3. Total Scores for M-CHAT-R and M-CHAT-R/F-A by Screening Outcome

M-CHAT-R-A Total Scores	Mean (SD)
Total sample, *n*=2594	1.18 (1.96)
Screened Negative on M-CHAT-R-A, *n*=2341	0.74 (.73)
Screened Positive on M-CHAT-R-A, *n*=253	5.32 (3.95)
No Evaluation Needed, M-CHAT-R-A, *n*=101	4.05 (2.29)
Qualified for Evaluation, M-CHAT-R-A, *n*=26	11.15 (4.63)
Refused Evaluation, *n*=6	10.83 (4.26)

M-CHAT-R/F-A Total Scores	Mean (SD)
Completed follow-up, *n*=127	1.72 (3.72)
No Evaluation Needed, M-CHAT-R/F-A, *n*=101	0.05 (.22)
Qualified for Evaluation M-CHAT-R/F-A, *n*=26	8.23 (3.77)
Refused Evaluation, *n*=6	6.33 (4.32)

Evaluated Sample Total Scores	M-CHAT-R-A Mean (SD)	M-CHAT-R/F-A Mean (SD)
ASD Group (ASD+Autism, *n*=17)	11.71 (4.07)	9.18 (3.59)
ASD, *n*=3	13.33 (6.03)	6.33 (3.51)
Autism, *n*=14	11.36 (3.75)	9.79 (3.42)
Non-ASD, *n*=2	9.50 (2.12)	6.00 (2.83)

Table 4. Frequency of M-CHAT-R-A Total Scores and Screening Outcomes

M-CHAT-R-A Total Score	Continued to Screen Positive After M-CHAT-R/F-A (Evaluation Needed), *n*		Lost to Follow up, *n*	Total Screened Positive on M-CHAT-R-A, *n*	Total, *n*	Percentage of Children Who Continued to Screen Positive After M-CHAT-R/F-A
	No	Yes				
0	0	0	0	0	1005	0%
1	0	0	0	0	947	0%
2	0	0	0	0	389	0%
3	63	2	69	134	134	1%
4	22	0	19	41	41	0%
5	4	0	6	10	10	0%
6	3	2	10	15	15	13%
7	2	2	1	5	5	40%
8	1	4	2	7	7	57%
9	1	0	4	5	5	0%
10	1	2	1	4	4	50%
11	1	2	4	7	7	29%
12	0	1	1	2	2	50%
13	1	2	1	4	4	50%
14	1	3	1	5	5	60%
15	1	1	1	3	3	33%
16	0	1	0	1	1	100%
17	0	0	4	4	4	0%
18	0	3	0	3	3	100%
19	0	1	2	3	3	33%
Total	101	26	126	253	2594	1%

Table 5. M-CHAT-R/F-A Total Scores (Follow Up Interview)

	Needed Evaluation?	
M-CHAT-R/F-A Total Score	Yes	No
0	0	96
1	0	5
2	1	0
3	2	0
4	1	0
5	3	0
6	4	0
7	1	0
8	3	0
10	2	0
11	4	0
12	2	0
14	1	0
15	2	0
Total	26	101

Table 6. Demographic Characteristics of Children Evaluated

Variable	Total
Total Sample	20
Sex	
Male	15
Female	5
Maternal Education	
No formal education/incomplete primary school	0
Unfinished secondary school	6
Secondary education	7
Some college	0
College/University Degree	2
Advanced Degree	0
Unknown	5
Person who Completed M-CHAT-R-A	
Mother	20
Age at screening, mean (SD), mo	26.05 (5.30)
M-CHAT-R-A total score, mean (SD), *n*=20	11.25 (4.83)
M-CHAT-R/F-A total score, mean (SD), *n*=20	8.80 (3.50)

Table 7. Evaluation Data by Case

		2 Stage Screening Total Scores		ADOS Total Scores					PEDS Totals		PED-DM: Percent of Delay				
Client	ADOS Class	M-CHAT-R-A	M-CHAT-R/F-A	Comm	RSI	Comm & RSI	Play	RRBs	Pred concerns	Non-pred concerns	Fine Motor	Self Help	Recep Lang	Exp Lang	Gross Motor
1	Non-ASD	11	8	5	2	7	2	0	3	4	54%	58%	19%	42%	-4%
2	Autism	6	6	5	13	18	4	0	2	2	45%	41%	5%	45%	32%
3	Autism	11	11	6	13	19	4	3	3	3	-4%	30%	74%	48%	-4%
4	Autism	12	11	6	14	20	4	4	3	3	71%	62%	100%	62%	71%
5	ASD	7	6	4	6	10	3	5	2	3	0%	8%	38%	0%	-29%
6	ASD	14	10	3	12	15	4	4	2	3	35%	74%	83%	61%	35%
7	Autism	8	5	6	13	19	4	4	2	1	56%	48%	74%	85%	0%
8	Autism	10	11	4	14	18	3	4	2	5	33%	31%	75%	58%	14%
9	Autism	13	15	8	10	18	4	1	4	5	33%	41%	78%	56%	44%
10	Non-ASD	8	4	2	4	6	4	2	3	3	25%	25%	75%	67%	14%
11	Autism	14	11	7	12	19	4	4	3	3	38%	42%	83%	71%	25%
12	Autism	7	10	6	8	14	4	4	3	4	33%	39%	75%	58%	31%
13	Autism	18	5	4	12	16	4	0	3	3	66%	54%	74%	66%	29%
14	Autism	6	6	6	14	20	4	1	2	0	10%	30%	40%	40%	25%
15	Autism	15	12	4	13	17	4	2	3	3	65%	53%	82%	65%	44%
16	Autism	13	12	4	14	18	4	2	3	2	65%	53%	82%	65%	26%
17	ASD	19	3	4	5	9	4	1	3	2	12%	53%	74%	65%	9%
18	Autism	16	15	5	9	14	1	0	3	0	4%	36%	76%	52%	4%
19	Autism	10	7	8	8	16	3	0	3	1	11%	63%	22%	56%	0%
20	No ADOS	14	8	n/a	n/a	n/a	n/a	n/a	2	0	0%	19%	72%	53%	3%
Average (SD)				5.11 (1.59)	10.32 (3.80)	15.42 (4.38)	3.58 (0.84)	2.16 (1.77)	2.70 (0.57)	2.5 (1.50)	32.56%	42.95% (16.30%)	65.03% (25.53%)	55.69% (16.79%)	18.43% (22.39%)

Table 8. Comparison of Demographic Characteristics for Screened Positive Children: Conducted vs. Lost to Follow Up

Variable	M-CHAT-R/F-A Conducted	Refused M-CHAT-R/F-A or Unable to Contact
Total Screen Positive on M-CHAT-R-A	127	126
Sex		
Male	50	59
Female	77	67
Maternal Education		
No formal education/incomplete primary school	4	2
Unfinished secondary school	40	54
Secondary education	43	25
College/University Degree	27	23
Advanced Degree	1	1
Unknown	12	21
Person who Completed M-CHAT-R-A		
Mother	116	92
Father	8	27
Other Family Member		3
Unknown	3	4
Age at screening, mean (SD), mo	24.36 (3.98)	24.09 (3.65)
M-CHAT-R-A total score, mean (SD)	5.50 (4.09)	5.13 (3.82)

Table 9. Demographic Characteristic Comparisons: Refused Evaluation vs. Evaluated Toddlers

Variable	Evaluation Conducted	Refused Evaluation
Total Screen Positive on M-CHAT-R-A	20	6
Sex		
Male	15	4
Female	5	2
Maternal Education		
No formal education/incomplete primary school	0	0
Unfinished secondary school	6	4
Secondary education	7	1
College/University Degree	2	1
Advanced Degree	0	0
Unknown	5	0
Person who Completed M-CHAT-R-A		
Mother	19	5
Father	1	1
Other Family Member	0	0
Unknown	0	0
Age at screening, mean (SD), mo	26.2 (5.20)	23.67 (4.72)
M-CHAT-R-A total score, mean (SD)	11.25 (4.83)	10.83 (4.26)
M-CHAT-R/F-A total score, mean (SD)	8.80 (3.50)	6.33 (4.32)

Table 10. Response Stability from M-CHAT-R-A to M-CHAT-R/F-A

	M-CHAT-R-A						M-CHAT-R/F-A			
Item #	Passed	Failed	Changed after Follow up (Passed)	% Changed after Follow up	Missing	Total	Pass	Fail	#Pass of Missing	# Fail of Missing
1	103	23	10	43%	1	127	113	14	1	0
2**	82	43	42	98%	2	127	126	1	2	0
3	78	44	28	64%	5	127	109	18	5	0
4	109	17	13	76%	1	127	123	4	1	0
5**	62	61	54	89%	4	127	119	8	4	0
6	84	11	5	45%	32	127	114	13	31	1
7	86	37	19	51%	4	127	107	20	4	0
8	99	26	11	42%	2	127	109	18	2	0
9	87	36	18	50%	4	127	109	18	4	0
10	105	18	11	61%	4	127	119	8	4	0
11	110	15	8	53%	2	127	119	8	2	0
12**	40	85	83	98%	2	127	124	3	2	0
13	120	4	2	50%	3	127	125	2	3	0
14	106	18	9	50%	3	127	117	10	2	1
15	98	23	16	70%	6	127	117	10	5	1
16	88	35	19	54%	4	127	109	18	3	1
17	84	34	18	53%	9	127	109	18	8	1
18	96	28	17	61%	3	127	114	13	2	1
19	90	33	23	70%	4	127	115	12	3	1
20	114	11	8	73%	2	127	124	3	2	0

**Reverse scored items.

Table 11. Frequency of Missing Items

M-CHAT-R-A Item		Response on Screener		
#	Item	Missing	No	Yes
1	If you point at something across the room, does your child look at it?	4	41	2549
2	Have you ever wondered if your child was deaf?*	8	2416	170
3	Does your child play pretend or make-believe?	20	117	2457
4	Does your child like climbing on things?**	5	60	2529
5	Does your child make unusual finger movements near his or her eyes?*	15	2196	383
6	Does your child point with one finger to ask for something or to get help?	192	18	2384
7	Does your child point with one finger to show you something interesting?	10	79	2505
8	Is your child interested in other children?	5	52	2537
9	Does your child show you things by bringing them to you or holding them up for you to see – not to get help, but just to share?	11	66	2517
10	Does your child respond when you call his or her name?	13	26	2555
11	When you smile at your child, does he or she smile back at you?	7	27	2560
12	Does your child get upset by everyday noises?*	15	1368	1211
13	Does your child walk?**	10	24	2560
14	Does your child look you in the eye when you are talking to him or her, playing with him or her, or dressing him or her?	10	40	2544
15	Does your child try to copy what you do?	13	47	2534
16	If you turn your head to look at something, does your child look around to see what you are looking at?	13	86	2495
17	Does your child try to get you to watch him or her?	29	68	2497
18	Does your child understand when you tell him or her to do something?	9	49	2536
19	If something new happens, does your child look at your face to see how you feel about it?	9	67	2518
20	Does your child like movement activities?**	10	31	2553

* Reverse Scored Items; "YES" response indicates ASD symptom presence

** Indicates foil items

Table 12. DFA Results: M-CHAT-R-A Items in Function 1

Item #	Question	Function 1
1	If you point at something across the room, does your child look at it?	0.629
10	Does your child respond when you call his or her name?	0.525
11	When you smile at your child, does he or she smile back at you?	0.491
9	Does your child show you things by bringing them to you or holding them up for you to see – not to get help, but just to share?	0.472
18	Does your child understand when you tell him or her to do something?	0.448
16	If you turn your head to look at something, does your child look around to see what you are looking at?	0.427
17	Does your child try to get you to watch him or her?	0.413
19	If something new happens, does your child look at your face to see how you feel about it?	0.400
14	Does your child look you in the eye when you are talking to him or her, playing with him or her, or dressing him or her?	0.385
15	Does your child try to copy what you do?	0.360
8	Is your child interested in other children?	0.346
7	Does your child point with one finger to show you something interesting?	0.337

DFA Results: M-CHAT-R-A Items Not Included in Function

Item #	Question	Function 1
3	Does your child play pretend or make-believe?	0.222
2	Have you ever wondered if your child was deaf?*	0.150
4	Does your child like climbing on things?**	0.136
20	Does your child like movement activities?**	0.108
5	Does your child make unusual finger movements near his or her eyes?*	0.086
6	Does your child point with one finger to ask for something or to get help?	0.084
13	Does your child walk?**	0.037
12	Does your child get upset by everyday noises?*	-0.035

** Indicates Reverse scored item, e.g., "Yes" response indicates ASD symptom presence*
*** Indicates foil items*

Table 13. M-CHAT-R/F-A with Problematic Items Removed, a Comparison with Original

	Original M-CHAT-R-A		M-CHAT-R-A Without Three Reverse Scored Items	
Screening Category	*n*	Mean (SD)	*n*	Mean (SD)
M-CHAT-R-A	2594	1.18 (1.96)	2594	0.49 (1.68)
Screened Negative on M-CHAT-R-A	2341 (90%)	0.74 (0.73)	2487 (96%)	0.21 (0.47)
Screened Positive on M-CHAT-R-A	253	5.32 (3.95)	107	7.07 (4.29)
No Evaluation Needed, M-CHAT-R-A	101	4.05 (2.29)	35	4.91 (3.01)
Qualified for Evaluation, M-CHAT-R-A	26	11.15 (4.63)	24	10.42 (3.91)
Lost to Follow up/Unable to Contact	126	5.13 (3.82)	48	6.96 (4.28)
Refused Evaluation	6	10.83 (4.26)	4	10.75 (6.08)

Appendix B
Figures

Figure 1. Flow Chart

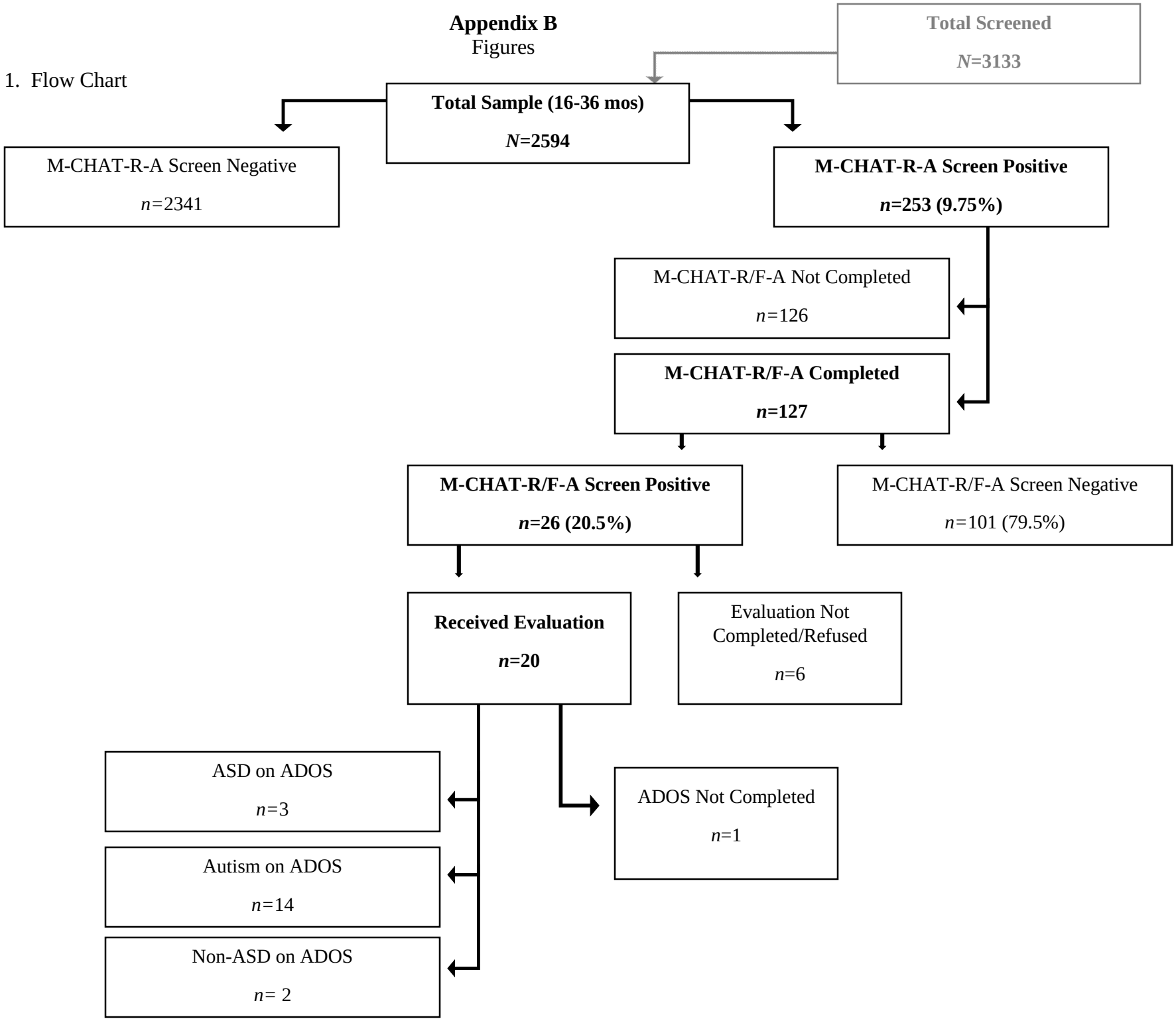

Figure 2. Revised Flow Chart