

# Validation non-English version of modified checklist for autism in toddlers-revised with follow-up

Karina Nukeshtayeva<sup>1</sup>, Marina Lubchenko<sup>2</sup>, Bauyrzhan Omarkulov<sup>3</sup>, Nailya Delellis<sup>4</sup>

<sup>1</sup>School of Public Health and Biomedicine, Karaganda Medical University, Karaganda city, Republic of Kazakhstan

<sup>2</sup>Department of Neurology, Neurosurgery, Psychiatry and Rehabilitology, Karaganda Medical University, Karaganda city, Republic of Kazakhstan

<sup>3</sup>Institute of Public and Occupational Health, Karaganda Medical University, Karaganda city, Republic of Kazakhstan

<sup>4</sup>Department of Health Science, The Herbert H. & Grace A. Dow College of Health Professions, Central Michigan University, Mount Pleasant, USA

Received: 2021-03-31.

Accepted: 2021-06-16



This work is licensed under a  
Creative Commons Attribution 4.0  
International License

J Clin Med Kaz 2021; 18(4):4-11

Corresponding author:

Karina Nukeshtayeva.

E-mail: [nukeshtaeva@qmu.kz](mailto:nukeshtaeva@qmu.kz);

ORCID: 0000-0002-4463-6874.

## Abstract

**Background:** In modern conditions, the diagnosis of autism spectrum disorders (ASD) in young children is quite complex. A two-step version of the updated checklist for autism in young children, revised with follow-up (M-CHAT-R/F), was shown to be successful in a large, geographically diverse sample of young children in primary care. In this study, the authors conducted a critical study of 16 validation studies of 35,722 participants participating in M-CHAT-R/F for early detection of ASD in non-English-speaking countries, and established a recommendation for validation of the non-English-language version of the M-CHAT-R/F screening method among toddlers.

**Aim:** Examine available publications on assessing the accuracy of the screening tool for detecting ASD in non-English speaking countries and identify possible errors in methodology.

**Material and methods:** The survey was conducted in databases. The criteria for inclusion of publications in this review: studies of the diagnostic accuracy of the screening tool for early detection of ASD M-CHAT-R/F, non-English speaking country of study. The search depth is 6 years.

**Results:** The sensitivity and accuracy of M-CHAT-R/F screening displayed outstanding performance with an average of 0.86 (CI=0.95) and 0.80 (CI=0.95) respectively. Based on the findings of this analysis, it can be inferred that, at some stage of the assessment of the validity and reliability of the M-CHAT-R/F instrument, each study included in the analysis rendered various degrees of error.

**Conclusion:** Non-compliance with the blindness guideline and the presence of children diagnosed with ASD and documented screening outcomes greatly raises the likelihood of error. A major challenge for researchers at the timing and flow stage was to ensure that the maximum number of participants participated from the very beginning to the end. It is more effective for researchers to work with a sample of small magnitude since this is an incentive to offer diagnostics to the full number of study participants.

**Key words:** validation, screening, autism spectrum disorder, cultural adaptation

## Introduction

A mixture of social communication disorders and repeated sensorimotor behaviors is autism spectrum disorder (ASD). Genetic and environmental influences are often correlated with ASD [1]. Research indicates that there is no widely recognized definition of the existence of ASD, and the lack of appropriate therapeutic options makes it necessary to consider ASD as an incurable condition at present [2]. Moreover, the proper diagnosis of ASD

requires daily evaluation and behavior review [3]. Also, research reports that individuals with ASD have difficulties in obtaining health services and lack of support throughout their lives [4,5].

In this respect, the value of detailed early detection is greatly enhanced. Initiated at the age of less than three years, intervention greatly enhances responsive behavior, raises IQ, and decreases the prevalence of ASD disease [6].

Even so, under current settings, it is very difficult to diagnose ASD under young children. Only 4 out of 21 cases of ASD have been reported by primary health care providers, suggesting the need for additional use of appropriate screening methods [7]. Minimum preparation and good outcomes with a higher positive prognostic value (PPV) are necessary for the use of screening methods by health professionals [8].

The early version of the Modified Checklist for Toddler Autism (M-CHAT) consists of 23 questions that were created in 1999, and has been proven in more than 22 countries to be accurate and effective. M-CHAT demonstrates outstanding diagnosis of small children with other developmental disabilities, but demonstrated poor PPV in a cohort of children under the age of 27 months [9].

The Modified Checklist for Autism in Toddlers-Revised Follow-Up (M-CHAT-R/F) was developed and tested in 2014 for the early diagnosis of ASD in children aged 16-30 months. The age of diagnosis of ASD can be lowered to 2 years by screening, thereby increasing the time available for early detection. Compared to the previous iteration of M-CHAT, M-CHAT-R/F detects ASD with a higher frequency and also displays slightly less false-positive findings owing to the second screening phase-additional interview [10]. The efficacy of the two-stage screening variant on a wide, geographically distributed primary care population of young children has been shown [11].

In terms of its usefulness, however, M-CHAT-R/F is a questionnaire focused on the child's parental assessment. Therefore, due to variations in parents' opinions on their children's normal actions, the efficacy of the screening method can vary [12]. Therefore, validity and reliability tests must be properly conducted in order to prevent overestimating or underestimating the usefulness of the M-CHAT-R/F models.

In this review, the authors carried out a critical examination of M-CHAT-R/F validation studies for early identification of ASD in toddlers. Analysis vulnerabilities and concerns that contributed to a high risk of testing errors were identified, as well as methods of addressing them.

## Material and methods

The criteria for inclusion in this review were: studies aimed at evaluating the validity and reliability of the M-CHAT-R/F screening tool in young children.

A literature research was done in common databases such as Scopus, Web of Science, eLibrary and Google Scholar. In total 775 references were found by the 'M-CHAT-R/F' keyword. Among them only 16 articles described validation trials of the M-CHAT-R/F screening instrument among children 24-30 months of age who are not at risk in a non-English - speaking country.

Final papers were evaluated using QUADAS-2 consistency evaluation diagnostic test tool [13]. Evaluation was conducted with a series of signaling questions, which were grouped into four categories: patient selection, index test, reference standard and timing and flow. Additionally, the evaluation questions were increased by following questions: training for health providers, training for the caregiver and cross-cultural adaptation. All signaling questions are listed in Table 2.

Applicability considerations are based on the following criteria developed to achieve the objective of the research, namely: the validation and reliability of the ASD M-CHAT-R/F early detection screening technique in non-English speaking countries. Exclusion criteria: studies conducted in an English-speaking country, evaluating the accuracy of the English-language version of M-CHAT-R/F, a study focused on the study

of risk groups for ASD, and not on the diagnostic accuracy of M-CHAT-R/F, using only earlier versions in the study questionnaires such as M-CHAT.

As indicated by the QUADAS-2 criteria, the degrees of uncertainty regarding the applicability of the analysis in the sample should be rated as "low," "high," or "unclear." It should be noted that the "unclear" word is used where there is insufficient data to assess the applicability of research in this review (Table 5). First of all, the applicability of the studies in this review was evaluated. In Table 5, can be seen the level of applicability concerns. Despite the fact that some studies have a high level of applicability concerns, all studies were included in the review because they have low-level applicability concerns in the "index-test" section.

The answers to each question could be 'yes', 'no' and 'unclear'. "Unclear" is written as an answer to a signal question when there is no information about this item in the publication.

## Results

A brief description of the studies included in the survey is provided in Table 1.

**Table 1** Description of the studies included in the review

Study	Sample size	Country	Year	Authors
1	2594	Albania	2016	{Formatting Citation}
2	120	Chili	2019	15
3	7928	China	2019	16
4	100	Egypt	2017	17,
5	6237	India	2018	18
6	110	Indonesia	2016	19
7	947	Mali	2019	20
8	163	Pakistan	2020	21
9	105	Russia	2018	22
10	1207	Saudi Arabia	2019	23
11	148	Serbia	2016	24
12	92	Slovakia	2017	25
13	6625	Spain	2018	26
14	369	Taiwan	2019	27,
15	9010	Turkey	2019	28,
16	27	Uzbekistan	2019	29.

## Patient selection

Answers on signaling questions are shown in Table 2.

10 of the studies (62.5%) avoided the "case-control" study design, 3 (18.75%) of the studies used it and there was no information to answer this question for the remaining 3 (18.75%) of the studies.

Consecutive or random approach of choosing children for testing was used in 10 of the studies (62.5%), 3 (18.75%) of the studies did not follow these recommendations and 3 (18.75%) of the studies did not provide any details about recruiting sample subjects.

Only 3 (18.75%) of the studies used acceptable inclusion and exclusion criteria. On the contrary, 8 (50%) of the studies included children with or at risk of diagnosis with ASD, behavioral illness and other neurodevelopmental conditions. Remaining 5 (31.25%) studies did not provide any details on the inclusion and exclusion requirements in the publications (Table 2).

Table 2

Answers on signaling questions

Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Patient selection																
Avoided case-control study design	Yes	No	yes	yes	Yes	yes	Unclear	unclear	No	Yes	no	yes	yes	Yes	yes	Unclear
A consecutive or random sample of patients enrolled	yes	Yes	Yes	unclear	yes	Yes	Unclear	yes	No	Yes	yes	No	no	Yes	Yes	Unclear
Avoided inappropriate exclusions	No	no	yes	No	unclear	yes	Unclear	unclear	no	no	no	no	No	Yes	unclear	unclear
Index-test																
M-CHAT-R/F results interpreted without knowledge of the results of diagnostic?	no	Yes	yes	No	yes	Yes	yes	Yes	No	yes	yes	no	no	Yes	yes	Unclear
If a threshold was used, was it pre-specified?	Yes	yes	Yes	Unclear	unclear	yes	Yes	yes	Yes	unclear	unclear	unclear	unclear	Yes	unclear	unclear
Reference standard																
Reference standard correctly identified ASD?	yes	Unclear	yes	Yes	unclear	Yes	unclear	yes	No	yes	0	No	yes	yes	Yes	no
Diagnostic results interpreted without knowledge of the results of the M-CHAT-R/F?	unclear	no	Unclear	no	no	yes	no	no	no	no	no	No	no	no	no	No
Timing and flow																
The appropriate interval between screening and reference evaluation	unclear	unclear	unclear	unclear	unclear	Yes	unclear	unclear	unclear	unclear	Yes	unclear	yes	Yes	unclear	unclear
All patients receive a diagnostic	no	no	No	yes	Yes	Yes	No	no	No	no	No	Yes	No	No	no	No
Patients receive the same reference standard	yes	Yes	No	Yes	unclear	Yes	Yes	No	yes	no	unclear	no	unclear	No	yes	unclear
All patients included in the analysis	no	no	no	Yes	Yes	Yes	no	No	No	No	no	yes	no	No	No	no
Training for health providers	Yes	Yes	unclear	unclear	unclear	Unclear	Yes	Unclear	unclear	yes	Unclear	yes	Yes	yes	Yes	unclear
Training for parents	Yes	unclear	Yes	Unclear	unclear	yes	Yes	unclear	Unclear	Yes	unclear	unclear	Yes	unclear	unclear	Unclear
Cross-cultural adaptation of M-CHAT-R/F	Yes	yes	yes	unclear	yes	Yes	yes	yes	yes	Yes	yes	no	yes	yes	Yes	no

## Index-test

The principle of blindness was found in 10 studies (62.5%), M-CHAT-R/F screening was conducted in children prior to the diagnostic test for ASD, and screening findings were also determined without impacting diagnostic results. However, as some research involved children with prior conditions, in 5 studies (31.25%) the theory of 'blindness' may not be possible (Table 2).

The threshold value of the original variant of M-CHAT-R/F for the possibility of ASD was tentatively used in half of the experiments (Table 2). These studies identified their thresholds, ranging from 2 to 7. No information about their thresholds was given for the second half of the studies (Table 3).

To prevent misinterpretation, it is very important to ensure that parents grasp the M-CHAT-R/F questions [30]. However, only a few studies offered short guidance for the doctors and caregivers involved in the study (8 out of 16 studies (50 %) and 6 out of 16 studies (37.5 %), respectively). However, no detail was contained in the other articles regarding any preparation or clarification (Table 2).

Cultural modifications to their versions of the M-CHAT-R/F questionnaire were carried out by the vast majority of publications (13 (81,2 %)). The two experiments did not carry out any cultural adaptation, but instead checked localized versions of the official website's M-CHAT-R/F questionnaire [31]. In its analysis, one study did not mention the reality of cultural adaptation. (Table 2)

The sensitivity and accuracy of M-CHAT-R/F screening displayed outstanding performance with an average of 0.86 (CI=0.95) and 0.80 (CI=0.95) respectively. In Study 7 (Sen=0.5) and in Study 9 (Sp=0.05), the minimum values are marked. Positive predictive value (PPV) and negative predictive value (NPV) mean values are 0.50 and 0.97, respectively. These metrics have not been measured in several tests (Table 3).

## Reference standard

In 14 (87.5 percent) trials, only children at high risk of ASD were diagnosed on the basis of the findings of the M-CHAT-R / F screening, which suggests that the diagnosis of the participants in the research was made with knowledge of the screening data, which is unacceptable for an independent examination of the questionnaire's validity and reliability (Table 1). Two studies (Study 5, Study 16) did not diagnose the study participants.

The diagnostic approaches used by the study authors are substantially different in Table 3. Most studies, however, have used the fifth edition of the Diagnostic and Statistical Manual of Psychiatric Disorders (DSM-5). The Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Program (ADOS) have the greatest database of evidence and the highest reliability and accuracy and were suggested for combined use [32]. However, this powerful mixture was only seen in one analysis (Study 12).

Table 3

Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood (LG+), negative likelihood (LG-), the cut-off value

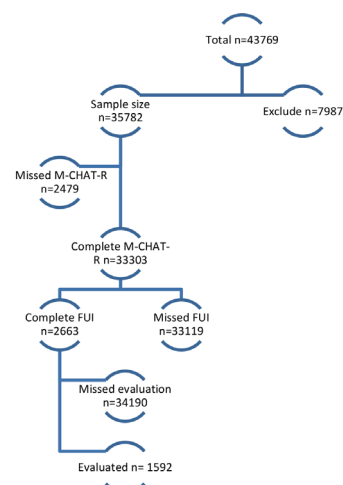
Study	Sensitivity	Specificity	PPV	NPV	LG-	LG+	Cut-off value
Study 1	-	-	0.14	-	-	-	5
Study 2	1	0.98	-	-	-	-	3
Study 3	0.96	0.85	0.69	1	7.12	0.042	3
Study 4	1	0.53	0.4	1	-	-	-
Study 5	-	-	-	-	-	-	-
Study 6	0.89	0.95	0.76	0.98	0.117	16.04	7
Study 7	0.5	1	1	0.87	0.5	-	2
Study 8	-	-	-	-	-	-	2
Study 9	0.83	0.05	-	-	-	-	2
Study 10	-	-	-	-	-	-	-
Study 11	-	-	-	-	-	-	-
Study 12	-	-	-	-	-	-	-
Study 13	0.79	0.99	0.39	0.99	0.21	215.99	-
Study 14	0.86	0.96	0.59	0.99	0.14	19.3	2
Study 15	1	0.91	0.086	1	-	-	-
Study 16	-	-	-	-	-	-	-

## Timing and flow

All the research analyzed included 35,782 individuals in total. Owing to non-compliance with the eligibility criterion (age not included in the 16-30-month group, diagnosis of ASD or other behavioral and developmental disabilities, lack of permission to enroll in the study), 7987 participants were omitted from the research. 2,479 participants were omitted because of inappropriate filling in the M-CHAT-R method of the first level of screening. There is a substantial decline in patient involvement (2663 participants) in the second stage of the screening due to comprehension issues, reluctance to further participate in the research, and communication issues. In studies, the cumulative number of individuals were diagnosed is just 1,592 (Figure 1).

Participants completed a screening test, Follow-Up, and diagnostics in 4 studies. In comparison, most experiments (12 out

Figure 1 - Number of participants at each stage of the study



of 16) did not involve all empirical participants. This is because most studies have involved people who were found to have a high to medium risk of ASD to undergo screening testing, while the study did not involve children with a moderate risk of ASD in diagnostics. Of the 16 reports, 12 did not have any detail on the period of time between screening and diagnosis, making it hard to determine its accuracy. And, as multiple diagnostic methods were used in the experiments, participants were exposed to multiple diagnostic methods in five studies (Table 2).

## Discussion

### Patient Selection

The presence of selective or random sampling of young children in research was examined in this segment, namely, the age group of 24-30 months that did not have an early registered diagnosis of "autism spectrum disorder," "Atypical autism," "Childhood autism," "Asperger syndrome," "Pervasive developmental disorder," and other neurodevelopmental, mental, and behavioral conditions. It is also mandatory that the experiments take into consideration the adequacy of the caregivers, their emotional and physical state. This is relevant because the parents of the children under review carry out M-CHAT-R/F screening, and inaccurate outcomes can lead to an over-optimistic or negative estimation of the efficacy of M-CHAT-R/F. Studies should also include both children and parents who satisfy the requirements stated and lack any grounds for exclusion. (Signaling questions: "A consecutive or random sample of patients enrolled", "Avoided case-control study design")

The "case-control" study design should be avoided if children are afflicted with ASD or other neurodevelopmental, mental, behavioral disorders to assess the efficacy of M-CHAT-R/F in a high-risk population. Since this design distorts fact and will lead to a re-evaluation of the M-CHAT-R/F screening tool's usefulness and accuracy (Signaling question: "Avoided case-control study design")

Children who had already been diagnosed with ASD or other mental and developmental disorders at the time of study were included in the failed trials. Children with ASD as a "case" group and children without diagnosis as a "control" category were used in some trials, so the "case-control" model method was used.

There's a strong chance of prejudice for this block in half of the tests. This is attributed largely to inadequate requirements for inclusion and exclusion. There is a low chance in 6 (37.5%) studies and there is no evidence on the procedures of recruiting of sample subjects and the conditions for participation and exclusion in 3 (18.8%) studies (Table 5).

### Index-test

In this analysis, M-CHAT-R/F is the index-test. The lack of diagnostic test findings concerning the outcomes of the M-CHAT-R/F screening should be the main issue of this segment. (Signalling question: 'M-CHAT-R/F findings perceived without information of test outcomes') In this question, owing to the arbitrary analysis of the M-CHAT-R/R questions, the concept of 'blindness' must be followed. First of all, for caregivers interested in the analysis, this idea should be followed. Until screening, health care providers should stop advising or posing questions about the progress of the child and mental health, as well as including parents with prematurely diagnosed children with ASD, other mental, neurodevelopmental, and behavioral

disabilities [33].

In addition, despite the fact that the initial version of M-CHAT-R / F sets cut-off scores for low, medium, and high risk, all children participating in the research must undergo follow-up interviews and diagnostic testing. Each analysis should measure the sensitivity and specificity of each "bad" item after passing all stages of screening and diagnostics and determine its cut-off values. (Signalling question: "Was it pre-specified whether a threshold was used"). Only in each analysis should the cut-off values of the initial M-CHAT-R / F version be suggested, but not final, as this could lead to an overestimation of the efficacy of M-CHAT-R / F, which could be lower in another sample of children with the same threshold score [13,34].

This analysis also looked at whether there was a short description of M-CHAT-R / F screening for caregivers and health professionals included in the research in primary health settings. (Signaling questions: "Health professional training", "Caregiver training"). This is important because this test is a trial in most trials because it has not historically been observed by any caregivers or health professionals. It is also important to clarify the purpose of the screening, its understanding, to the health care professionals engaging in the research, so that they can correctly communicate it to caregivers. For the most reliable and relevant screening, it is therefore important to communicate with clinicians about the purposes and value of screening and receive informed consent from them. For the same purposes, the cultural elements of the M-CHAT-R/F screening questions need to be updated. For both health care professionals and caregivers, for proper understanding and clarification of screening questions. (Signaling question: «Cross-cultural adaptation of M-CHAT-R/F»)

Most studies had a low risk of error for this block (10 (62,5%), 5 (31,3%) studies had a high risk, and 1 (6,3%) study had inadequate risk assessment information. In experiments with a high probability of error (Table 5), the key issue being that children with an older diagnosis were included in the studies, which violated the 'blindness' concept in the study of the efficacy of M-CHAT-R/F.

### Reference standard

This section analyzes the approaches used in trials to identify ASD in toddlers. It is believed that the diagnostic approaches used will assess ASD by 100% (Signaling query "Reference form correctly defined ASD"), and inaccurate screening outcomes are the discrepancies between the diagnostic results and the results of M-CHAT-R/F [35].

Around the same time, the idea of 'blindness' must be followed, as in the previous section. Knowing the effects of screening will influence the outcomes of diagnostics and lead to a re-evaluation of the efficacy and detection accuracy of ASD using M-CHAT-R/F. (Signaling question: «Diagnostic findings perceived without knowledge of the effects of the M-CHAT-R/F») [33,36].

Most studies (11 (68.8 %) have a high probability of error for this block, since the studies interpreted the testing findings for those members of the sample that had a high and average probability of ASD, suggesting that the screening outcomes were known. For this block, only Study 1 (6.3 %) has a low chance of a mistake (Table 5).

### Flow and timing

Ideally, it is important to quickly process the findings of screening and diagnostic testing. As ASD is marked by

**Table 4** Diagnostic methods used in studies

Study	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6	Study 7	Study 8	Study 9	Study 10	Study 11	Study 12	Study 13	Study 14	Study 15	Study 16
Evaluation type	ADOS	ADOS-2	CARS DSM-5	CARS	NA	DSM-5	Multidisciplinary	CARS-2 DSM-5	ADOS	DSM-5 Multidisciplinary	Multidisciplinary	ADOS-2 ADI-R	ADOS-2 DSM-5	DSM-5 Multidisciplinary	ADOS-2 DSM-5	NA

**Table 5** Risk of bias in studies evaluating the validity and reliability of the tool, as well as applicability concerns to review the question of screening for the presence of ASD M-CHAT-R/F

Study	Risk of bias				Applicability concerns		
	Patient selection	Index-test	Diagnostics	Flow and timing	Patient selection	Index-test	Diagnostics
Study 1	High	low	unclear	High	low	low	low
Study 2	High	low	high	High	high	Low	low
Study 3	low	low	unclear	High	low	low	low
Study 4	high	unclear	High	Low	High	low	Low
Study 5	low	low	High	unclear	Unclear	low	high
Study 6	low	low	low	Low	Low	low	low
Study 7	unclear	low	high	High	unclear	Low	low
Study 8	unclear	low	high	High	low	low	Low
Study 9	high	high	high	High	High	low	low
Study 10	low	low	high	High	low	low	Low
Study 11	high	high	unclear	High	High	Low	low
Study 12	high	high	high	High	High	low	Low
Study 13	high	low	high	High	High	Low	low
Study 14	Low	High	high	High	low	low	Low
Study 15	low	low	high	High	low	Low	low
Study 16	unclear	high	High	High	unclear	Low	high

a persistent path, however, no regression or regeneration is expected. So a couple of days' delay does not seem like a concern. Around the same time, for all research subjects, the longer the distance between screening and diagnosis, the greater the chance of low diagnostic coverage. This would also, in fact, lead to an erroneous estimation of the efficacy of M-CHAT-R/F screening in studies. ("Appropriate period between screening and comparison assessment" signaling question)

To prevent bias, experiments do not use multiple testing approaches under the influence of screening outcomes. As there are many screening approaches, this is particularly important for the diagnosis of ASD. It is therefore important to diagnose all research subjects, irrespective of the outcomes of M-CHAT-R/F screening, to prevent inappropriate sensitivity and specificity evaluation. Both children examined should be included in the study, regardless of the ASD screening and diagnostic findings [37]. (Signaling questions: «All patients receive a diagnostic», «All patients receive a diagnostic», «All patients included in the analysis»).

If the knowledge given in the analysis satisfies the above criteria, the answer to the signal question is "yes". If, on the other hand, it does not satisfy, then the answer to the query of the signal is written as 'no.' The response to the signal question is written as "unclear" if there is inadequate knowledge in the analysis on the questions posed.

Just two studies have a low risk of error when testing the efficacy of their variant of the M-CHAT-R/F for this block (Table 4). The majority of experiments have a high probability of error, largely because not all subjects in the sample have completed a diagnostic test and have not been included in the final review.

### Conclusion

Based on the findings of this analysis, it can be inferred that, at some stage of the assessment of the validity and reliability of the M-CHAT-R/F instrument, each study included in the analysis rendered various degrees of error. The researchers had the largest number of problems, evidently at the "Reference Standard," "Timing and Flow" level. This may be because there is little consensus in the question of the "gold standard" in diagnosis of ASD in realistic health care (Table 5). And this can be seen in Table 4, where we can see that various diagnostic approaches and techniques are used from sample to research. Around the same time, non-compliance with the blindness guideline and the presence of children diagnosed with ASD and with documented screening outcomes greatly raises the likelihood of error. Future studies to determine the validity and efficacy of the M-CHAT-R/F screening method in its community should also concentrate on these two topics. Next, instrumental diagnostic approaches for ASD, such as ADOS-2 and ADI-R, can also be used for

diagnostic purposes in combination with the multidisciplinary diagnosis of infants. Second, eliminate any chances to include children with current ASD diagnoses and other developmental disabilities in trials to prevent prejudice.

A major challenge for researchers at the timing and flow stage was to ensure that the maximum number of participants participated from the very beginning to the end. However, as can be seen in Figure 1, there is a substantial decline in the number of participants at the diagnostic level. This is attributed to a number of causes. From an excessively large initial number of participants participating in the study to a low level of health literacy in the study population or problems with health promotion activities. Thus, in future studies, at this point, researchers should pay more attention to the sample under analysis and set an acceptable timeline between each stage of the study. Attention should also be given to increasing the trust of the sample subjects in the study themselves and to carrying out explanatory practices.

Study 6 is the study with the lowest error chance (Table 5). After evaluating all the products, it can be inferred that this is partly due to the comparatively limited sample size (110). It is as straightforward as possible for such a sample to go through all phases of the analysis. Around the same time, it is more effective for researchers to work with a sample of this magnitude since this is an incentive to offer diagnostics to the full number of study participants.

In addition to the aforementioned guidelines on sample size, the following relevant points are recommended for prospective researchers:

- participation in the report by people who have not previously had an ASD diagnosis and other psychiatric and developmental disorders.
- the age of the children tested should be between 16 and 30 months.
- the lack of motor and speech disabilities in the sample participants;
- the implementation of the "gold standard" in the diagnosis of ASD.
- maximum participation of all participants in the research process at each stage;
- cultural adaptation of the M-CHAT-R/F version.

**Disclosures:** There is no conflict of interest for all authors.

**Acknowledgements:** None

**Funding:** None

## References

1. Lord C, Elsabbagh M, Baird G, Veenstra-Vanderweele J. Autism spectrum disorder. *Lancet*. 2018; 392(10146):508-520. [https://doi.org/10.1016/S0140-6736\(18\)31129-2](https://doi.org/10.1016/S0140-6736(18)31129-2)
2. Lavrov N V, Shabanov PD. Autism spectrum disorders: etiology, treatment. models and experimental studies. *Rev Clin Pharmacol Drug Ther*. 2018; 16(1):21-27. <https://doi.org/10.17816/rcf16121-27>
3. Filipek PA, Accardo PJ, Baranek GT, et al. The screening and diagnosis of autistic spectrum disorders. In: *The Science of Mental Health: Volume 2: Autism*. Taylor and Francis; 2013; 11-56. doi:10.1201/9780203026229.ch6
4. King C, Merrick H, Le Couteur A. How should we support young people with ASD and mental health problems as they navigate the transition to adult life including access to adult healthcare services. *Epidemiol Psychiatr Sci*. 2020; 29(e90):1-5. <https://doi.org/10.1017/S2045796019000830>
5. An S, Chan CK, Kaukenova B. Families in Transition: Parental Perspectives of Support and Services for Children with Autism in Kazakhstan. *International Journal of Disability, Development and Education*. 2018; 67(1):28-44. <https://doi.org/10.1080/1034912X.2018.1499879>
6. Dawson G, Rogers S, Munson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: *The early start Denver model*. *Pediatrics*. 2010; 125(1):e17-e23. <https://doi.org/10.1542/peds.2009-0958>
7. Robins DL. Screening for autism spectrum disorders in primary care settings. *Autism*. 2008; 12(5):537-556. <https://doi.org/10.1177/1362361308094502>
8. Turner ML, Wendy LS, Pozdol SJ, et al. Follow-up of children with autism spectrum disorders from age 2 to age 9. *Autism*. 2006; 10(3):243-265. <https://doi.org/10.1177/1362361306063296>
9. Toh TH, Tan VWY, Lau PST, Kiyu A. Accuracy of Modified Checklist for Autism in Toddlers (M-CHAT) in Detecting Autism and Other Developmental Disorders in Community Clinics. *J Autism Dev Disord*. 2018; 48(1):28-35. <https://doi.org/10.1007/s10803-017-3287-x>
10. Robins DL, Casagrande K, Barton M, Chen CMA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*. 2014; 133(1):37-45. <https://doi.org/10.1542/peds.2013-1813>
11. Chlebowski C, Robins DL, Barton ML, Fein D. Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*. 2013; 131(4):e1121-7. <https://doi.org/10.1542/peds.2012-1525>
12. Zachor D, Yang JW, Itzchak E Ben, et al. Cross-cultural differences in comorbid symptoms of children with autism spectrum disorders: An international examination between Israel, South Korea, the United Kingdom and the United States of America. *Dev Neurorehabil*. 2011; 14(4):215-220. <https://doi.org/10.3109/17518423.2011.568468>
13. Whiting P, Rutjes A, Westwood M, Mallett S. Updating QUADAS: Evidence to inform the development of QUADAS-2. *BristolAcUk*. 2010. <http://www.bristol.ac.uk/quadas/resources/quadas2reportv4.pdf>.
14. Brennan L, Fein D, Como A, Rathwell IC, Chen CM. Use of the Modified Checklist for Autism, Revised with Follow Up-Albanian to Screen for ASD in Albania. *J Autism Dev Disord*. 2016; 46(11):3392-3407. <https://doi.org/10.1007/s10803-016-2875-5>
15. Coelho-Medeiros ME, Bronstein J, Aedo K, et al. Validación del M-CHAT-R/F como instrumento de tamizaje para detección precoz en niños con trastorno del espectro autista. *Rev Chil pediatría*. 2019; 90(5):492-499. <https://doi.org/10.32641/rchped.v90i5.703>

16. Guo C, Luo M, Wang X, et al. Reliability and Validity of the Chinese Version of Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F). *J Autism Dev Disord.* 2019; 49(1):185-196. <https://doi.org/10.1007/s10803-018-3682-y>
17. El Mougy MT, Ahmed MHN, Hassan HA. Early Detection of Autistic Spectrum Disorder (ASD) by Modified Checklist for Autism in Toddlers, Revised (M-Chat-R) in Children with Social and Verbal Delay. *Al-Azhar Med J.* 2017; 46(1): 71-78. <https://doi.org/10.12816/0035536>
18. TS J, Jacob P, Srinath S, et al. Toddlers at risk for Autism Spectrum Disorders from Kerala, India – A community based screening. *Asian J Psychiatr.* 2018; 31:10-12. <https://doi.org/10.1016/j.ajp.2017.12.016>
19. Windiani IGAT, Soetjningsih S, Adnyana IGAS, Lestari KA. Indonesian Modified Checklist for Autism in Toddler, Revised with Follow-Up (M-CHAT-R/F) for Autism Screening in Children at Sanglah General Hospital, Bali-Indonesia. *Bali Med J.* 2016; 5(2):133. <https://doi.org/10.15562/bmj.v5i2.240>
20. Sangare M, Toure HB, Toure A, et al. Validation of two parent-reported autism spectrum disorders screening tools M-CHAT-R and SCQ in Bamako, Mali. *eNeurologicalSci.* 2019; 15:100188. <https://doi.org/10.1016/j.ensci.2019.100188>
21. Farooq A, Ahmed S. Screening for Autism Spectrum Disorder in Children up to Age 2.5 years in a Tertiary Care Hospital. *Life Sci.* 2020; 1(1):11-15. <https://doi.org/10.37185/Ins.1.1.74>
22. Pervushina ON, Kudrjavceva AS, Mezenceva MP. Adaptacija oprosnika Modified checklist for autism in toddlers, revised with Follow-Up (M-CHAT-R/F) dlja skrininga detej rannego vozrasta. *Psihologicheskij instrumentarij.* 2018; 11(1):47-67
23. Alawami AH, Perrin EC, Sakai C. Implementation of M-CHAT Screening for Autism in Primary Care in Saudi Arabia. *Glob Pediatr Heal.* 2019; 6:1-6. <https://doi.org/10.1177/2333794x19852021>
24. Carakovac M, Jovanovic J, Kalanj M, et al. Serbian Language version of the Modified Checklist for Autism in Toddlers, Revised, with Follow-Up: Cross-Cultural Adaptation and Assessment of Reliability. *Sci Rep.* 2016; 6:7-11. <https://doi.org/10.1038/srep38222>
25. Hnilicová S, Celušaková H, Hnilica P, Babinská K, Pivovaráčiová A, Ostatníková D. Screening for autism spectrum disorders in population of young children in Slovakia. *Act Nerv Super Rediviva.* 2017; 59(1):29-32.
26. Magán-Maganto M, Canal-Bedia R, Hernández-Fabián A, et al. Spanish Cultural Validation of the Modified Checklist for Autism in Toddlers, Revised. *J Autism Dev Disord.* 2018; 50(7):2412-2423. <https://doi.org/10.1007/s10803-018-3777-5>
27. Tsai JM, Lu L, Jeng SF, et al. Validation of the modified checklist for autism in toddlers, revised with follow-up in Taiwanese toddlers. *Res Dev Disabil.* 2019; 85:205-216. <https://doi.org/10.1016/j.ridd.2018.11.011>
28. Oner O, Munir KM. Modified Checklist for Autism in Toddlers Revised (MCHAT-R/F) in an Urban Metropolitan Sample of Young Children in Turkey. *J Autism Dev Disord.* 2019; 50:3312-3319. <https://doi.org/10.1007/s10803-019-04160-4>
29. Gavrilova TA, Madzhidova YN, Mukhammadsolikh SB, Khusenova NT. SCREENING DIAGNOSTICS M-CHAT-R FOR IDENTIFYING AUTHISTIC SPECTRUM DISORDERS. 2019; 2(1):33-38.
30. Herlihy L. Racial/ethnic and Socioeconomic Differences in Screening Toddlers for Autism Spectrum Disorders Using the M-CHAT. Dr Diss. June 2014. <https://opencommons.uconn.edu/dissertations/454>. Accessed February 21, 2020.
31. M-CHATM - MCHAT R/F Translations. <https://mchatscreen.com/mchat-rf/translations/>. Accessed May 13, 2020.
32. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: A systematic literature review. *Eur Child Adolesc Psychiatry.* 2013; 22(6):329-340. <https://doi.org/10.1007/s00787-013-0375-0>
33. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of Variation and Bias in Studies of Diagnostic Accuracy: A Systematic Review. *Ann Intern Med.* 2004; 140(3):189-202. <https://doi.org/10.7326/0003-4819-140-3-200402030-00010>
34. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: Mechanisms, magnitude, and solutions. *Clin Chem.* 2008; 54(4):729-737. <https://doi.org/10.1373/clinchem.2007.096032>
35. Biesheuvel C, Irwig L, Bossuyt P. Observed differences in diagnostic test accuracy between patient subgroups: Is it real or due to reference standard misclassification? *Clin Chem.* 2007; 53(10):1725-1729. <https://doi.org/10.1373/clinchem.2007.087403>
36. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011; 155(8):529-536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
37. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Chapter 10 Analysing and Presenting Results. <http://srdta.cochrane.org/>. Accessed May 7, 2020.