Journal of Autism and Developmental Disorders (2020) 50:4218–4220
https://doi.org/10.1007/s10803-020-04462-y

LETTER TO THE EDITOR

Please, Don't Shoot the Meta-analysis: A Response to "A Commentary to Toddler Screening for Autism Spectrum Disorder: A Meta-analysis of Diagnostic Accuracy by Sánchez-Garcia et al. 2019"

Ana B. Sánchez-García[1] · Ana B. Nieto-Libreros[2] · Purificación Galindo-Villardón[2] · Diana L. Robins[3]

Published online: 25 March 2020

In their letter to the editor, Øien et al. (2019) highlight concerns about conducting a meta-analysis on screening instruments for ASD. They focus on lack of information and heterogeneity among the individual studies pooled in the research by Sánchez-García et al. (2019). The authors do not question the experimental method; even stating that it "*provides us with important knowledge of how screening instruments perform across various studies.*" Of course, clinical and methodological heterogeneity could cause statistical heterogeneity leading to inaccurate conclusions in a meta-analysis (Higgins and Altman 2008). Although this should not preclude conducting meta-analytic investigation, since understanding the causes of heterogeneity increases its scientific value and the clinical significance of the results (Thompson 1994). Therefore, it is not a threat or concern to synthesize the available evidence (Lijmer et al. 2002). Having said that, the methodological limitations among studies included in a meta-analysis is not specific to ASD screening; yet, there is a large literature about the viability and importance of diagnostic or screening accuracy meta-analyses. Given empirical support in other disorders, it is logical to apply them to a meta-analysis in screening instruments for ASD. There are three specific points that warrant further consideration.

Heterogeneity in Autism Screening Tools Studies

Although one might wonder about the limitations of combining heterogeneous studies, heterogeneity is an asset rather than a problem to the modern meta-analysis (Arends 2006), because it allows us to investigate how potential sources of heterogeneity affect the psychometric properties of screening tools. The heterogeneity may be due to multiple causes e.g.: differences in cut-offs given that a population-specific calibration of the test to certain level of sensitivity will result in different cutoff values for different populations (Rücker et al. 2018); different index test and variation in threshold due to study design, patients and disease cohorts (Rutter and Gatsonis 2001); internal bias caused by methodological flaws (Jones et al. 2019); or technical aspects of the screenings tools and particular characteristics across studies (Rhodes et al. 2018; Macaskill et al. 2010). Likewise, some studies do not follow up on screen negative cases, which may affect sensitivity estimation. In this sense, we refer the interested reader to the work of Robins (2020) who reviews the scenarios that can lead to these false negatives and the advantages and disadvantages of the different approaches to detect them (concurrent, prospective, longitudinal). Moreover, in diagnostic or screening studies a negative correlation between sensitivity and specificity is expected because the cut-point used for a positive test result varies between studies (Macaskill 2004). So, all of these causes can be quantified using meta-analytic methods.

✉ Ana B. Sánchez-García
asg@usal.es

Ana B. Nieto-Libreros
ananieto@usal.es

Purificación Galindo-Villardón
pgalindo@usal.es

Diana L. Robins
dlr76@drexel.edu

[1] INICO- University of Salamanca, Pso. Canalejas, 169, 37008 Salamanca, Spain

[2] Department of Statistics, University of Salamanca, 37008 Patio de Escuelas, Spain

[3] AJ Drexel Autism Institute, 3141 Chestnut Street, Philadelphia, PA 19104, USA

Modelling the Heterogeneity in Autism Screening Tools

A meta-analysis of screening test accuracy must identify and summarize evidence on the accuracy of tests, including an assessment of the quality, consistency of the evidence and heterogeneity. According to Takwoingi et al. (2015) pooling the results of multiple studies provides a more precise estimate of test performance than a single study. The Bayesian Hierarchical Model employed in our study is robust in adjusting for the imperfect nature of autism screenings tools given that it is statistically rigorous to model the within-study binomial structure of the data and to accounting for between-study heterogeneity in TP and FP rates (Rutter and Gatsonis 2001). In fact, HSROC model considers that the relationships between sensitivities and specificities are non-linear; hence, it is not necessary to check baseline data to ensure the accuracy of the analysis, because it takes into account the correlation between the observed sensitivity and 1-Specificity for each study (Rutter and Gatsonis 2001). The aforementioned not only supports the sampling variability but it allows summary estimates of sensitivity and specificity (Macaskill 2004). We also delved into understanding the heterogeneity among the included papers to know the confidence intervals that describe the relationship between the psychometric properties. Although traditional ROC curves plot sensitivity against 1-specificity, the HSROC meta-analysis model instead plots Sensitivity against the False Positive rate. This allows us to evaluate how variability in cutoff value affects the heterogeneity of diagnostic accuracy data. In this way, we see which studies are responsible for high levels of heterogeneity, how cut-off values varied, and how moderate negative correlations between sensitivities and False Positive rates are.

We agree there are a wide variety of studies with different characteristics that makes not all of them equal; *the difficulty lies is deciding just how similar studies need to be to be integrated into a meta-analysis of screenings tools*. This is one of the reasons why it is necessary to define inclusion criteria to identify studies directly address the review question, and to subject the primary studies to stringent quality control analysis. In order to produce a set of studies that could be combined for the meta-analysis, Sánchez-García et al. (2019) clearly specified the inclusion and exclusion criteria. In addition, the researchers carried out a rigorous data quality analysis, and including a heterogeneity study. As the heterogeneity was high, thus they used a random effects model (DerSimonian and Laird 1986) to describe the variability in test accuracy across studies and carried out a subgroup analysis. Quality control measures including QUADAS quality analysis, publication bias and subgroup analysis were performed and we removed from analysis studies that did not provide sufficient data to construct a 2×2 contingency table (such as those without confirmatory diagnoses for screen positive cases) or those, which had a low quality rating in the quality assessment (see Sánchez-García et al. 2019). Thus, it is possible to combine different studies to obtain a synthesis measure, provided that the process is carried out with sufficient rigor.

Qualitative Reviews or Meta-analysis

Finally, in the last paragraphs of the letter, the authors say "*we are more inclined to follow conclusions by for instance the National Institute for Health and Care Excellence (NICE) in the UK and argue that we are not there yet (Baird* et al. 2011*; UK National Screening Committee* 2012)". The works that they reference, while relevant, are based on qualitative reviews, which provide descriptive evidence, and are methodologically very different from the quantitative meta-analysis of Sánchez-Garcia et al. (2019). The preference for non-statistical methodologies rather than quantitative integration of multiple studies neglects the advantage of meta-analyses, which increases "*power to detect real differences in test accuracy between tests than single studies, and may yield more precise estimates of expected sensitivity and specificity*…" (Macaskill et al. 2010, p.4).

Although we agree that the state of science for ASD screening is imperfect and will benefit from continued study, we believe that the meta-analysis results inform clinical and health policy decision making (Rutter and Gatsonis 2001). In short, whether we want it or not, the meta-analysis is just the *messenger* that visualizes some of the challenges presented by the research field. So that, *Please don´t shoots the Meta-analysis of diagnostic accuracy in autism*.

Acknowledgements We are grateful that the authors of the letter have allowed a scientific space to discuss and address aspects that could raise doubts, since ultimately one of the functions of a meta-analysis is to generate a scientific debate in the research area.

References

Arends, L. R. (2006). Multivariate Meta-analysis: Modelling the Heterogeneity. PhD diss., Erasmus University, Rotterdam, Netherlands.

Baird, G., Douglas, H. R., & Murphy, M. S. (2011). Recognising and diagnosing autism in children and young people: Summary of NICE guidance. *BMJ*, *343*, d6360. https://doi.org/10.1136/bmj.d6360.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188.

Higgins, J. P., & Altman, D. G. (2008). *Assessing risk of bias in included studies. Cochrane handbook for systematic reviews of*

interventions: Cochrane book series (pp. 187–241). Hoboken: Wiley.

Jones, H. E., Gatsonsis, C. A., Trikalinos, T. A., Welton, N. J., & Ades, A. E. (2019). Quantifying how diagnostic test accuracy depends on threshold in a meta-analysis. *Statistics in Medicine, 38*, 4789–4803.

Lijmer, J. G., Bossuyt, P. M., & Heisterkamp, S. H. (2002). Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in Medicine, 21*(11), 1525–1537.

Macaskill, P. (2004). Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology, 57*(9), 925–932.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). *Cochrane handbook for systematic reviews of diagnostic test accuracy. Version 0.9.0*. London: The Cochrane Collaboration.

Øien, R. A., Cicchetti, D. V., Nordahl-Hansen, A., & Schjølberg, S. (2019). A commentary to "toddler screening for autism spectrum disorder: A meta-analysis of diagnostic accuracy". *Journal of Autism and Developmental Disorders, 49*, 1837–1852.

Rhodes, K. M., Turner, R. M., Savović, J., Jones, H. E., Mawdsley, D., & Higgins, J. P. (2018). Between-trial heterogeneity in meta-analyses may be partially explained by reported design characteristics. *Journal of Clinical Epidemiology, 95*, 45–54.

Robins, D. L. (2020). How do we determine the utility of screening tools? *Autism, 24*(2), 271–273.

Rücker, G., Steinhauser, S., Kolampally, S., & Schwarzer, G. (2018). Diagmeta: Meta-analysis of diagnostic accuracy studies with several cutpoints.

Rutter, C. M., & Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine, 20*(19), 2865–2884.

Sánchez-García, A. B., Galindo-Villardón, P., Nieto-Librero, A. B., Martín-Rodero, H., & Robins, D. L. (2019). Toddler screening for autism spectrum disorder: A meta-analysis of diagnostic accuracy. *Journal of Autism and Developmental Disorders, 49*(5), 1837–1852.

Takwoingi, Y., Riley, R. D., & Deeks, J. J. (2015). Meta-analysis of diagnostic accuracy studies in mental health. *Evidence-Based Mental Health, 18*(4), 103–109.

Thompson, S. G. (1994). Systematic review: Why sources of heterogeneity in meta-analysis should be investigated. *BMJ, 309*(6965), 1351–1355.

UK National Screening Committee. (2012). Screening for autistic spectrum disorders in children under the age of five policy position statement and summary. Retrieved August 22, 2019, from https://legacyscreening.phe.org.uk/autism.