Quality measures of two-stage newborn hearing screening: Systematic review with a Bayesian meta-analysis.

Short title: Meta-analysis of two-stage newborn hearing screening

Kirsi Manz[1*], Uta Nennstiel[2], Carola Marzi[2], Ulrich Mansmann[1¶], Inken Brockow[2¶]

[1]Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Faculty of Medicine, LMU Munich, Munich, Germany

[2]Bavarian Health and Food Safety Authority, Oberschleißheim, Germany

*Corresponding author

E-mail: manz@ibe.med.uni-muenchen.de (KM)

[¶]These authors contributed equally to this work.

Abstract

Background: Newborn Screening for hearing impairment (NHS) is a crucial public health issue worldwide. Often, a two-stage screening with two different testing approaches is used. We aimed to investigate the optimal screening algorithm, based on data from the literature published in the last 30 years. A particular focus of the study was to synthesize the existing evidence on two-stage newborn hearing screening regarding the refer rate (RFR), the percentage of children that did not pass the second test or were lost after the first test.

Methods: We searched MEDLINE and Scopus for studies on two-stage NHS using transitory evoked otoacoustic emissions (TEOAE) or automated auditory brainstem response (AABR). All studies on newborns who received their first test as an inpatient and a second test up to one month later were eligible. Random effects meta-analysis and Bayesian modeling were performed to estimate RFR, effects of the second test phase on the RFR, and sensitivity and specificity of TEOAE and AABR, respectively. Risk of bias was assessed using QUADAS-II. The unfunded study was registered in PROSPERO (CRD42023403091).

Results: Eighty-five study protocols, including over 1,12 million newborns, met the inclusion criteria. Certainty in the evidence was rated as moderate. The RFR was higher when the test method was changed than without a change of method (AABR-AABR: RFR = 1.3% (95% confidence interval (CI): 0.9, 1.8%), TEOAE-TEOAE: RFR = 2.7% (CI: 2.2, 3.2%), TEOAE-AABR: RFR = 3.9% (CI: 2.9, 5.1%), AABR-TEOAE: 5.9% (CI: 5.0, 6.9%). Across all protocols, both methods demonstrated very high sensitivity (AABR: 99.9% (95% credibility interval (CrI): 99.6, 100.0%), TEOAE: 100.0% (CrI: 99.9, 100.0%)) and high specificity (AABR: 96.9% (CrI: 96.8, 97.0%), TEOAE: 91.1% (CrI: 91.0, 91.2%)).

Conclusions: Strategies that did not involve changes to the screening method had lower RFR. Although both methods demonstrated high sensitivity, AABR appears to have slightly higher specificity compared to TEOAE.

Introduction

A properly functioning auditory system is essential for a child's acquisition of spoken language. Early intervention is critical for age-appropriate spoken language development in children with hearing impairment [1-5]. Therefore, the Joint Committee on Infant Hearing recommends that interventions begin no later than three months of age [6]. Since the late 1990s, two screening methods have been available which allow a very early diagnosis: the measurement of transitory evoked otoacoustic emissions (TEOAE) and automated auditory brainstem response (AABR) [7]. With these testing methods, a prevalence of about 1,3 per 1,000 newborns [8], good treatment options with hearing aids, cochlear implants and early intervention, and a positive cost-benefit balance [9-11] congenital hearing disorders are suitable as a target disease for newborn screening [12,13]. Universal newborn hearing screening (UNHS) was included in standard care in Germany in 2009, as specified by the Federal Joint Committee (G-BA) in §§ 47 to 57 of the Pediatrics Directive [14]. The Pediatrics Directive outlines a two-step screening algorithm in well babies, beginning with initial TEOAE measurement in both ears, followed by bilateral AABR measurement if the initial test is not passed, and defines quality criteria. An important quality factor of a screening program is the refer rate (RFR). In the NHS, this refers to the percentage of screened infants who require

referral to a pediatric audiologist for further diagnostic examinations after not passing the screening tests (positive screening). This includes newborns who are lost-to-follow-up after not passing the initial test, as well as those who receive a “fail” result on the second test. . In the Pediatrics Directive the RFR should not exceed 4 %. False positive findings are of particular concern as they can cause anxiety for affected families and increase the demand for scarce resources in pediatric audiology practices and outpatient clinics.

Evaluations of the German NHS for the years 2011/12 [8] and 2017/18 [15,16] have shown that the recommended screening-algorithm is often not followed. In more than 50% of cases where infants do not pass the first test, a second TEOAE measurement is performed instead of the required AABR. Additionally, this second TEOAE test yielded a “fail” result in only about 10% of cases, compared to 20% for infants who underwent a second test using AABR. The failure rate of the second test was particularly high when the screening method was altered. Analysis of these data demonstrated that the second test showed the lowest failure rate with the TEOAE / TEOAE algorithm at 9.62%, while the highest rate was observed with TEOAE / AABR at 26.59%.

Accordingly, international recommendations suggest performing a second TEOAE test after the initial TEOAE result was “fail” in newborns without risk factors for hearing impairment (“well-babies”) [6,17].

This study reviewed the current literature to investigate the quality of available screening tests and the optimal screening algorithm based on data published in the last 30 years. Specifically, the study focused on synthesizing existing evidence related to RFR in two-stage NHS.

Materials and methods

Model

The first step was to formalize the two-stage screening process on which our meta-analysis was based.

The study population consists of newborns with relevant hearing impairment (“diseased”, D^+) and those without (“healthy”, D^-). The prevalence of hearing impairment in the population is denoted by π. The first stage of screening is performed using a method with a sensitivity of SE1 and specificity of SP1. Therefore, the test positivity rate of the first stage (PR1), which includes true positives (TP1) and false positives (FP1), is given by: $PR1 = TP1 + FP1 = \pi*SE1 + (1-\pi)*(1-SP1)$. These newborns are considered “*failed”* and should receive a second test. The rate of positive results from the first stage is also referred to as the “failure rate”.

The proportion of newborns whose first test result was “pass” and who therefore leave the screening process is the negative rate $NR1 = \pi*(1-SE1) + (1-\pi)*SP1$. However, not all newborns with a positive first test proceed to the second stage: The proportion ρ of positively tested newborns is *lost*. This loss is assumed to be independent of hearing status (newborns drop out for reasons unrelated to the first test result). Thus, a proportion of $(1-\rho)$ newborns undergo the second stage screening test.

The proportion of SE2 is identified as positive among the true positives of the first test. Accounting for the loss, this yields TP2 = TP1*(1-ρ)*SE2 = π*SE1*(1-ρ)*SE2. Similarly, the proportion of (1-SP2) is falsely identified as positive among the false positives of the first test. Accounting for the loss, we obtain FP2 = FP1*(1-ρ)*(1-SP2) = (1-π)*(1-SP1)*(1-ρ)*(1-SP2).

The proportion of newborns with "fail" results in both tests, together with those who are lost after a first positive test, form the refer rate (RFR):

$$RFR = \rho*[\pi*SE1 + (1-\pi)*(1-SP1)] + (1-\rho)*[\pi*SE1*SE2 + (1-\pi)*(1-SP1)*(1-SP2)]. \quad \text{(Eq. 1)}$$

The structure of the two-stage model, together with the respective selection and loss to follow-up processes is shown in S1 Fig.

Theoretical consideration of factors influencing the refer rate
We consider the relationship between the first stage positive rate (PR1 = π*SE1 + (1-π)*(1-SP1)) and the refer rate (RFR).

$$\begin{gathered} RFR = \rho*[\pi*SE1 + (1-\pi)*(1-SP1)] + (1-\rho)*[\pi*SE1*SE2 + (1-\pi)*(1-SP1)*(1-SP2)] = \\ \rho*PR1 + (1-\rho)*[\pi*SE1*SE2 + (1-\pi)*(1-SP1)*(1-SP2)] = \\ PR1 * \{\rho + (1-\rho) * (\pi*SE1*SE2 + (1-\pi)*(1-SP1)*(1-SP2)) / (\pi*SE1 + (1-\pi)*(1-SP1))\} = \\ PR1* \{\rho + (1-\rho)*P(T2+|T1+) \}. \quad \text{(Eq. 2)} \end{gathered}$$

In case of a small prevalence π (close to 0, inserting π=0 in Eq.2), it holds

$$RFR = PR1 * \{\rho + (1-\rho)*(1-SP2) \}. \quad \text{(Eq. 3)}$$

For small prevalence estimates, the RFR is linearly related to the failure (positive) rate of the first stage PR1. This linear relationship is determined by the loss rate ρ and the specificity of the second stage test SP2. Such a linear relationship between the RFR and failure rate can be exploited in a meta-regression. The quantity P(T2+|T1+) is called the conditional stage 2 positive rate (cPR2).

Equations 2 and 3 split the RFR into a component that quantifies the influence of the first stage and a component that quantifies the influence of the procedures after stage 1 (selection to stage 2 and quality of the second test). The RFR is the product of these two factors. It is of interest to quantify influence after stage 1. This will result from a meta-regression.

Simultaneous estimation of sensitivity and specificity in different two-stage screening settings
For the simultaneous estimation of the sensitivity and specificity, we assume that sensitivities and specificities of each test method are equal across the studies under consideration. For a study-specific prevalence π_S and the two-stage procedure of the screening process (with equal sensitivities and specificities across all studies) the following holds:

Positive rate at stage 1: PR1 = P(T1+) = π_S*SE1 + (1-π_S)*(1-SP1).

Conditional positive rate at stage 2:

$$\begin{gathered} cPR2 = P(T2+|T1+) = \\ (\pi_S*SE1*SE2 + (1-\pi_S)*(1-SP1)*(1-SP2)) / (\pi_S*SE1 + (1-\pi_S)*(1-SP1)). \end{gathered}$$

The value of cPR2 is not influenced by the loss rate ρ.

Both positive rates can be estimated from the data. If multiple studies with different prevalences are available for each test combination used (TEOAE-TEOAE, TEOAE-AABR, AABR-TEOAE, AABR-AABR), it is possible to infer the corresponding sensitivities and specificities (see [18,19]). It is also possible to estimate the specific loss rate of the study protocol (how many test positives of the first stage are examined in the second stage) and the RFR (see Eq. 2).

Study protocol

The meta-analysis was prospectively registered in PROSPERO (CRD42023403091). The review protocol can be found at and downloaded from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023403091. No amendments to the protocol have been made.

Eligibility criteria

For this review, the following PICO criteria were applied:

(P) Population: (Well) babies undergoing a two-stage hearing screening: (1) Initial screening as an inpatient in the maternity clinic; (2) Second test up to a maximum of one month later; (3) No use of the test method distortion product otoacoustic emissions (DPOAE); (4) Not exclusively newborns from NICU (neonatal intensive care unit).

(I) Intervention: Two-stage hearing screening using TEOAE, AABR, or a combination of both

(C) Comparator: not applicable

(O) Outcome: RFR after two screening steps, sensitivity and specificity of TEOAE and AABR

The population should preferably consist of well babies. Studies that included newborns with risk factors for hearing impairment or babies from the NICU were included if the study data did not clearly distinguish between well babies and these newborns. However, studies that included only newborns with risk factors or from the NICU were excluded from this review.

The review considered studies in which both the first and the second tests were completed within one month after birth.

Search method

We searched MEDLINE using PubMed and Scopus for relevant articles without language or geographic restrictions from the time of their inception through February 9th, 2024. The following search strategy was used for both databases:

("newborn hearing screening") OR ("neonatal hearing screening") OR ("infant hearing screening") OR ((("newborn screening") OR ("neonatal screening") OR ("infant screening")) AND ("hearing")) OR (("hearing screening") AND ("newborn" OR "neonatal" OR "infant"))

A broader search strategy without specification of test method (TEOAE or AABR) was chosen to avoid missing relevant publications that may not specify the test method in the title or abstract. This decision was based on a small pilot search conducted to test the search

strategy, which showed that including the type of screening test in the search resulted in missing some articles that were already known to be relevant to the review.

All articles had an abstract in English. If the full texts of the selected articles were not written in a language that the authors speak, online translation tools (DeepL, Google Translator) were used to translate them into English where possible. Only peer-reviewed publications were included.

Study records

The search strategy was saved in Citavi version 6.11. The data from the selected publications were extracted into Excel spreadsheets.

One reviewer searched the information sources and screened the titles and abstracts of the identified studies for inclusion and classified each study as eligible or ineligible. The study was classified as potentially eligible if it could not be clearly excluded based on its title and abstract. The full texts of all (potentially) eligible studies were then retrieved and reviewed by two additional reviewers. Again, studies were marked as eligible or ineligible for inclusion and the selection was discussed with the first reviewer until all three reviewers agreed. The first reviewer extracted the required data and made the preliminary decision on study inclusion based on the availability of the data for extraction. The second reviewer double-checked the extracted data and made corrections, which were discussed with all three reviewers and led to the final inclusion of all studies with available data.

Data extraction and items

Data extracted for the study description included the following: First author and year of publication, screening test combination (i.e., TEOAE-TEOAE, TEOAE-AABR, AABR-TEOAE, AABR-AABR), years of screening, country, number of newborns screened, name of screening device(s), time of first test and second test, and whether the study was conducted only in well babies or also in newborns with risk factors/ from the NICU.

Data extracted for quantitative analysis included the following: Number of newborns screened with the first test and type of first test (TEOAE/AABR), number of newborns who passed the first test, number of newborns who did not pass the first test, number of newborns who did not pass the first test and did not return for the second test (lost-to-follow-up), number of newborns screened for the second time and type of second test (TEOAE/AABR), number of newborns passing the second test, number of newborns who did not pass the second test, and number of newborns referred after not passing both screening tests.

Studies providing data on simultaneous TEOAE and AABR testing were added to both screening test combinations: TEOAE followed by AABR and AABR followed by TEOAE. Otherwise, such studies would have had to be excluded as the order of the tests was not sequential.

The following data were derived from the variables collected: The failure rate after the first test step was calculated as the quotient of the number of newborns who did not pass the first test and all newborns screened in the first step. Similarly, the failure rate after the second test step was calculated as the quotient of the number of newborns who did not pass the second test and all newborns screened in the second step. The RFR was calculated as the sum of the number of newborns who did not pass the second test and the number of newborns who did

not pass the first test and did not return for the second test, divided by the total number of newborns screened.

Outcomes and prioritization

Data were sought to calculate the RFR after the two-step NHS, with the failure rate after the first test included in the analysis. For the analysis, it was essential to obtain the number of infants who passed and did not pass each screening step. Therefore, studies that only reported the results after both screening steps could not be included in the review. This review focused mainly on well babies, but studies conducted in both well babies and newborns with risk factors or from the NICU were also included.

Risk of bias in individual studies and publication bias

The quality of the included studies was assessed using a modified version of the QUADAS-II tool developed for diagnostic accuracy studies [20]. Of the original four domains, the "reference test" domain was not applicable to our study because we focused on the two-stage screening process without knowing the true hearing loss status. Therefore, we replaced the domain "index test" by "first test" and "reference test" by "second test". The risk of bias for each included study was assessed independently by two reviewers. Disagreements were resolved by discussion with a mediator. The risk of bias was assessed at the study level.

Publication bias was not expected to have a significant impact on the literature found. It was assumed that all studies on two-stage NHS would be worthy of publication, as they describe not only the quality of the NHS, but also its implementation and problems. Selective reporting within studies, e. g. favoring one of the two screening tests, is also unlikely to be a problem. Therefore, methods to assess the risk of publication bias were not used in this meta-analysis.

Confidence in cumulative evidence

The strength of the overall body of evidence was assessed using Grading of Recommendations, Assessment, Development and Evaluation [21].

Subgroup analyses

Subgroup analyses were not specified in the study protocol. However, based on the characteristics of the included studies, we decided to analyze only well-baby studies as a subgroup. An additional post hoc sensitivity analysis was performed to assess the effect of outliers in the TEOAE-TEOAE group.

Statistical analysis

Continuous variables were summarized by median (minimum - maximum) or presented as box plots, and categorical variables were presented by frequency (%). Due to the exploratory nature of our study, adjustment for multiple testing was not considered. Statistical significance was claimed at 5% level ($p<0.05$) or for non-overlapping 95% credibility intervals. Random-effects meta-analysis of the RFR was performed using the R package *rmeta* and the DerSimonian-Laird approach. Heterogeneity indices Q and I^2 were calculated using the random effect estimates and random effect weights. For the visualization of the risk of bias, we used the source code of the *rob_summary* function of the *robvis* package to generate a similar graph adapted to our needs. Calculations were performed using WinBUGS version 1.4.3 [22] and R Version 3.6.3 [23]. All data and analysis scripts are available in the Open Science Framework repository at https://osf.io/nuk4p/.

Results

The PRISMA flow diagram for the search and study selection process is shown in Fig 1. Out of the 5886 records identified (PubMed: n=3356, Scopus: n=2530), 2000 duplicates were removed prior to screening. From a total of 3886 records screened, a total of 3563 records were excluded because the titles and abstracts of these articles were not relevant to our research question. A full-text search was conducted on the remaining 323 records. Seven records could not be retrieved. Out of the 316 reports that were screened, 239 were excluded. In many cases this was due to a failure to provide the required data at each screening step (n=77), an unsuitable study design (n=50), or not performing the first or second test within the specified time frame of one month (n=44). Additional reasons for exclusion are listed in Fig 1. N=7 reports comprised more than one screening protocol (n=6 two protocols [24-29] and n=1 three protocols [30]). Of the reports with more than one protocol, three [27,29,30] provided data on simultaneous TEOAE and AABR testing and were included in both the TEOAE-AABR and AABR-TEOAE groups. A total of 77 reports with 85 study protocols were included in the meta-analysis.

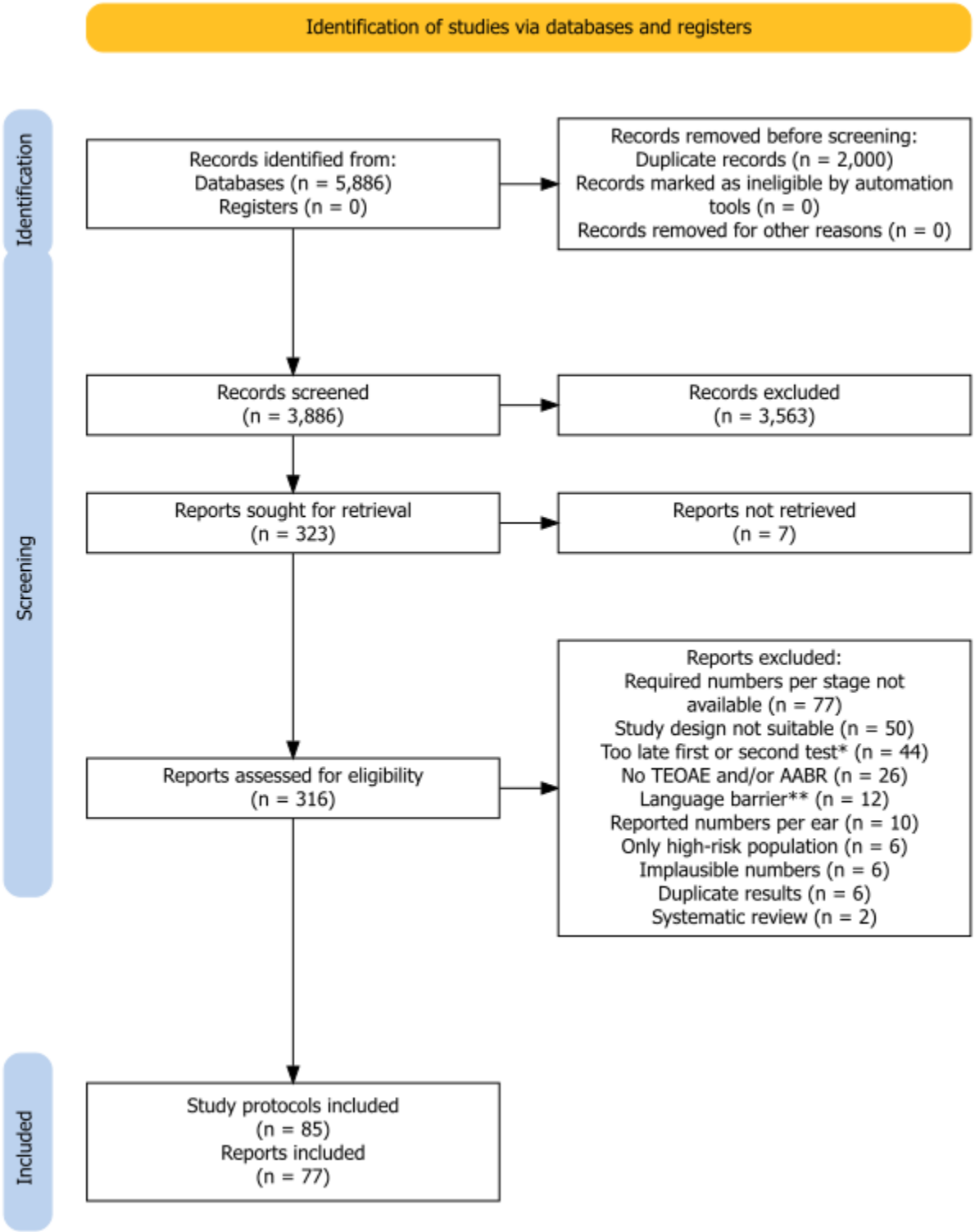


Fig 1. PRISMA flow diagram of the search and study selection process.
*Includes also studies where the time point of the first or second test was not specified. **Language barrier refers to reports that could not be automatically translated into English online. A detailed list of the 239 excluded reports is available from the authors upon request.

The analysis included 85 study protocols from 77 reports (7 of those with more than one study protocol) with 1125617 newborns. Table 1 provides an overview of the study characteristics. Of the 85 study protocols, n = 55 (64.7%) studies examined the TEOAE-TEOAE test

combination, n = 9 (10.6%) examined the TEOAE-AABR test combination n = 3 (3.5%) examined the AABR-TEOAE test combination, and n = 18 (21.2%) examined the AABR-AABR test combination. The median study size across all study protocols was n = 3238 newborns (min - max: 64 - 245219). The median study size for TEOAE-TEOAE was n = 3724 newborns (min - max: 81 - 245219), for TEOAE-AABR it was n = 3540 newborns (min - max: 64 - 50633), for AABR-TEOAE it was n = 247 newborns (min - max: 64 - 2005), and for AABR-AABR it was n = 3614 newborns (min - max: 81 - 199034).

Table 1: Characteristics of the included study protocols

Autor	**Year(s)**	**Country**	**No of screened newborns**	**Device 1st and 2nd test**	**Time of first test**	**Time of second test**	**High-risk/ NICU included***
TEOAE - TEOAE:							
Aidan et al. 1999 [31]	1995-1997	France	1421	ILO92	48 h	within the first month	no
Alanazi 2020 [32]	2020[J,D]	Saudi Arabia	20171	GSI AudioScreener	before hospital discharge	within two weeks	yes
Arjmandi et al. 2012 [33]	2009-2010	Iran	1232	unknown	10 days[E]	after 2-3 weeks	yes
Arora et al. 2022 [34]	2017-2019	India	1200	Neuro-Audio	< 72 h	after 3-4 weeks	yes
Arslan et al. 2013 [35]	2007-2008	Turkey	2229	Accuscreen	within 7 d, before hospital discharge	within 15 days after 1st screening	yes
Azizi et al. 2016 [36]	2006-2007	Iran	3818	Otoread	<48 h	after 2-4 weeks	yes
Benito-Orejas et al. 2008 [24]	2001-2003	Spain	2454	ILO92	within 48h, NICU: infants tested on discharge day	within the first month	yes
Bevilacqua et al. 2010 [37]	2010[J,I]	Brazil	11466	Audix	24 h	within 20 days after birth	yes
Busse et al. 2023 [25]	2018-2019	Albania	778	unknown	24-48 h	after 14 days	no
Calevo et al. 2007 (a) [38]	2001	Italy	3238	Echo Screen	48-72 h	within the third week of life	yes
Calevo et al. 2007 [39]	2002-2004	Italy	32502	Echo Screen	48-72 h	within the third week of life	no
Cavalcanti et al. 2012 [40]	2007-2009	Brazil	3724	AccuScreen	36-48 h	one week after first test	no[F]
Chapchap et al. 2001 [41]	1996-1999	Brazil	4196	ILO88	48-72 h (NICU: prior discharge)	within 30 days (no specification)	yes
Clarke et al. 2003 [26]	2001-2002	United Kingdom	81	Echocheck (ILO88)	21 h[A]	before hospital discharge	no
De Capua et al. 2007 [42]	1998-2006	Italy	19700	ILO 292	96 h[A], NICU: at post-menstrual age of 37-41 weeks	10-20 days after birth	yes
Diego et al. 2023 [43]	2016-2019	Brazil	1553	unknown	before hospital discharge	within 30 days	yes
Eibenstein et al. 2014 [44]	2007-2012	Italy	4579	AccuScreen	before hospital discharge	within 2 weeks	no
Eibenstein et al. 2015 [45]	2013-2014	Italy	3120	AccuScreen	before hospital discharge	within 2 weeks	no

Autor	Year(s)	Country	No of screened newborns	Device 1st and 2nd test	Time of first test	Time of second test	High-risk/ NICU included*
Erturk et al. 2010 [28]	2002-2003	Turkey	500	Echocheck, Accuscreen Pro	before hospital discharge	after 3 weeks	yes
Escobar-Ipuz et al. 2019 [46]	2007-2017	Spain	9350	unknown	<48 h	before hospital discharge	yes
Farahani et al. 2017 [47]	2013	Iran	2784	AccuScreen	day 1 or day 2	1-2 weeks after first test	no
Ferlito et al. 2021 [48]	2018	Italy	37562	unknown	48-72 h	within the first month	yes
George et al. 2016 [49]	2015	Bahrain	1834	unknown	before hospital discharge	1 week after discharge	yes
Fusetti et al. 2007 [50]	2005-2007	Italy	1400	Eclypse	within 1 week of life	2 weeks after first test	no
Gilbey et al. 2013 [51]	2010-2011	Israel	4958	unknown	before hospital discharge	day after 1st screening	no
Guastini et al. 2010 [52]	2006-2009	Italy	8671	ILO92	48-72 h, NICU: when stable general condition	two weeks after first test	yes
Gül et al. 2013 [53]	2010-2011	Tirkey	2363	Echoport ILO 288	before hospital discharge	two weeks after first test	yes
Habib et al. 2005 [54]	1996-2004	Saudi Arabia	11986	ILO88	<48 h	5th day	no
Hatzopoulus et al. 2007 [55]	2003-2004	Albania	450	EchoLab Plus	24-48 h	within 4 weeks of first test	no
Jakubikova et al. 2003 [56]	2003[J]	Sloval Republic	3048	ILO88	4th-12th day at hospital discharge	1 month after first test	yes
Kayiran et al. 2009 [57]	2004-2009	Turkey	8052	Accuscreen	before hospital discharge	1 week later	no
Konukseven et al. 2010 [58]	2007-2009	Turkey	1917	Ero-Scan	<48 h	after 10 days	no
Korres et al. 2006 (1) [59]	2006[J,G]	Greece	22195	ILO88	48-72h	after 1 month	yes
Korres et al. 2006 (2) [60]	2006[J,H]	Greece	25032	ILO88	48-72h	after 1 month	yes
Kosmidou et al. 2021 [61]	2018 - 2020	Greece	1491	Accuscreen	first days of life	within the first month	no
Lin et al. 2004 [62]	2000-2002	Taiwan	5938	EchoScreen	>24, before discharge	1 month later	no
Lotfi et al. 2007 [63]	2002-2004	Iran	7718	Echo-Screen	3-36 h	15-30 days old	no
Magnani et al. 2015 [64]	2010-2013	Italy	10359	AccuScreen	24-48 h	within 3weeks from birth	no
Molini et al. 2016 [65]	2010-2012	Italy	18796	unknown	24-36 h	1 month of age	no
Molteni 2006 [66]	1999-2005	Italy	10454	ILO OAE V5	3rd day of life	1 month	yes
Pastorino et al. 2005 [67]	1997-2001	Italy	19290	ILO88, Echoscreen	36-48 h (vaginal delivery), 3-5 d (C-section)	15 - 30 d after discharge	no
Pedersen et al. 2008 [68]	2006	Denmark	1627	ABaer	2-30 days	2-30 days	yes
Prpic et al. 2007 [69]	2002-2006	Croatia	11746	AuDX	2. or 3. day (NICU: when stable)	3 weeks after first test	yes
Pyarali et al. 2023 [70]	2021	Pakistan	267	GSI Screener	before hospital discharge	before hospital discharge	yes

Autor	Year(s)	Country	No of screened newborns	Device 1st and 2nd test	Time of first test	Time of second test	High-risk/ NICU included*
Satish et al. 2019 [71]	2015-2017	India	26487	unknown	<48 h	one week after first test	yes
Sennaroglu et al. 2011 [72]	2009-2010	Turkey	1840	AuDX-Pro	before hospital discharge/within 10 days	after 15 days	yes
Sequi Canet et al. 2016 [73]	2002-2013	Spain	14015	Echocheck ILO88	as late as possible prior discharge	age of 1 month	yes
Sheng et al. 2021 [30]	2018-2019	China	1340	AccuScreen	<48 h	<48 h	no
Tasci et al. 2010 [74]	2007-2008	Turkey	15323	AccuScreen	24-48 h	within 10 days	no
Tatli et al. 2007 [75]	2002-2003	Turkey	711	Echocheck	last working day prior hospital discharge	one week after first test	yes
Unlu et al. 2015 [76]	2009-2013	Turkey	2933	Echoport ILO292	day 5	day 15	no
Vaid et al. 2009 [77]	2005-2007	India	1238	unknown	< 72 h	1 month	no
Vos et al. 2014 [78]	2007-2012	Belgium	245219	Accuscreen, Echo-Screen, Echocheck	48-72 h	day 3 or 4 (following day after first test)	no
Welzl-Müller et al. 1997 [79]	1997[J]	Austria	2338	Echoport	within the first days	within 1-2 days after first test	no
Yorulmaz et al. 2017 [80]	2011-2016	Turkey	13693	Echo-Screen	before hospital discharge	14 days later	yes
TEOAE - AABR:							
Dort et al. 2000 [27]	2000[J]	Canada	64	ILO88, Smart Screener	before hospital discharge	before hospital discharge	yes
Lin et al. 2007 [81]	2004-2005	Taiwan	3540	ILO292, ALGO III	>48 h	before hospital discharge	yes
Mazlan et al. 2022 [82]	2010-2019	Malaysia	50633	unknown	<24 h	within 2 weeks	yes
Nennstiel-Ratzel et al. 2007 [83]	2003-2005	Germany	16767	Echo-screen, ALGO portable	before hospital discharge	before hospital discharge	yes
Olusanya et al. 2008 [84]	2005-2006	Nigeria	1150	Echo-Screen, ALGO Portable	24-48 h	before hospital discharge	no
Olusanya et al. 2010 [85]	2005-2007	Nigeria	4718	Echo-Screen, ALGO Portable	24 h, SCBU[B]: shortly before discharge	before hospital discharge	yes
Ong et al. 2020 [29]	2018	Philippines	247	Ero-Scan, AccuScreen	before hospital discharge	before hospital discharge	yes
Pasha et al. 2018 [86]	2006-2014	Iran	40930	Ero Scan, Eclipse EP25	before hospital discharge	1 month of age	no
Sheng et al. 2021 [30]	2018-2019	China	2005	AccuScreen, AccuScreen	< 72 h	< 72 h	no
AABR-TEOAE:							
Dort et al. 2000 [27]	2000[J]	Canada	64	ILO 88, Smart Screener	before hospital discharge	before hospital discharge	yes
Ong et al. 2020 [29]	2018	Philippines	247	Ero-Scan, AccuScreen	before hospital discharge	before hospital discharge	yes
Sheng et al. 2021 [30]	2018-2019	China	2005	AccuScreen, AccuScreen	< 72 h	< 72 h	no

Autor	Year(s)	Country	No of screened newborns	Device 1st and 2nd test	Time of first test	Time of second test	High-risk/ NICU included*
AABR - AABR:							
Alothman et al. 2024 [87]	2021	Saudi Arabia	199034	ALGO	24 h / prior discharge	prior discharge	yes
Al Shamisi et al. 2023 [88]	2010-2019	United Arab Emirates	37661	Integrity V500 System	before hospital discharge	1 month	yes
Ayas et al. 2021 [89]	2017-2020	United Arab Emirates	1821	Algo 3i	24-48 h or shortly before discharge	two weeks after first test	no
Benito-Orejas et al. 2008 [24]	2004-2006	Spain	3117	AccuScreen Pro	within 48h, NICU: tested on discharge day	within 1 month after birth	yes
Busse et al. 2023 [25]	2018-2019	Albania	1129	unknown	24-48 h	after 14 days	yes
Clarke et al. 2003 [26]	2001-2002	United Kingdom	81	ALGO-3	24 h[A]	before hospital discharge	no
Clemens et al. 2001 [90]	1999-2000	United States	3142	Algo 2/Algo 2e	before hospital discharge	within 12-24 h after first test, prior discharge	no
Erturk et al. 2010 [28]	2002-2003	Turkey	500	Accuscreen Pro	before hospital discharge	After 3 weeks	yes
Fan et al. 2010 [91]	2005-2008	Taiwan	7139	ALGO 3i	before hospital discharge	at 1 month of age	no
Gupta et al. 2015 [92]	2011-2012	India	2265	Beraphone MB11	24-48 h, preterm babies > 34 postmenstrual weeks	within 7 days after 1st test	yes
Huang et al. 2013 [93]	2009-2010	Taiwan	15790	unknown	24-36 h	36 - 60 h of age	yes
Iwasaki et al. 2003 [94]	2000-2001	Japan	4085	ALGO 2e	48-72 h	5 - 6 days after birth, prior discharge	yes
Kelly et al. 2021 [95]	2014-2016	United States	31984	ALGO	6h (vaginal delivery), 12 h (C-section)	before hospital discharge	no
Messner et al. 2001 [96]	1998-1999	United States	5771	ALGO-II	< 24 h	before hospital discharge	no
Oruc et al. 2021 [97]	2018	Turkey	5399	Accuscreen	before hospital discharge	15 days later	yes
Shim et al. 2021 [98]	2005-2015	South Korea	3059	Abaer system	24 h	within 1 month	yes
Tanyeri Toker et al. 2023 [99]	2020-2021	Turkey	570	unknown	<72 h	7-15 days of age	no
Tsuchiya et al. 2006 [100]	1999-2004	Japan	8979	ALGO 2e	day 4	1 month of age	no

*Studies that include only healthy newborns or exclude newborns with risk factors were scored as "no", NICU: newborn intensive care unit, [A]median age at first test, [B]SCBU: Special care baby unit, [C]second half of the Wessex trial, [D]Data from three years were analyzed, [E]mean age at first test, [F]originally only well babies, newborns transferred to special care baby unit were counted as losses to follow-up, [G]Three years after program initiation, [H]Additional data from two additional years were presented, [I]during 3 years, [J]Unknown dates, the publication year of the article is given.

Fig 2A presents the loss rates of newborns who did not pass the first test and did not attend the second test for all test combinations. Depending on the study, loss rates of up to 72% were reported. Median loss rates were 14% for TEOAE-TEOAE, 5% for TEOAE-AABR, and zero for AABR-TEOAE and AABR-AABR. Fig 2B shows the RFR of all study protocols. Median RFRs of 2.5%, 4.8%, 6.5% and 1.1% were found for the TEOAE-TEOAE, TEOAE-AABR, AABR-TEOAE, and AABR-AABR combinations, respectively. All but three studies [26,53,80], reported RFR below 10%. Out of the 85 study protocols, 57 (67.1%) showed a RFR below 4%.

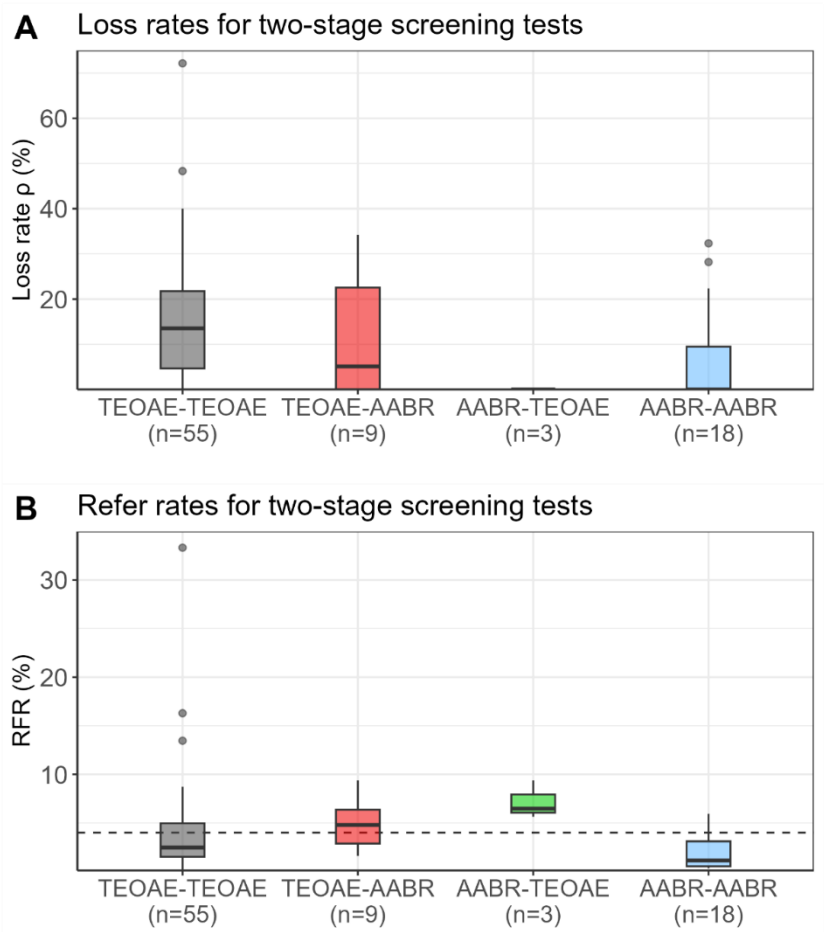


Fig 2. Loss rates (ρ) (A) and refer rates (RFR) (B) for the four different screening test combinations.

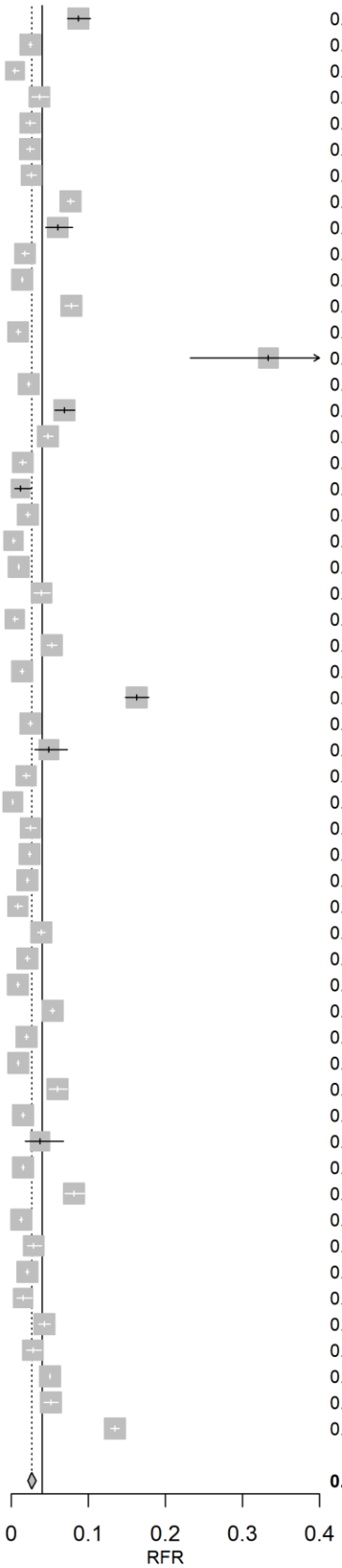
The dashed horizontal line indicates the 4% threshold quality criteria defined in the Pediatrics Directive for the RFR. AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission.

Random-effects meta-analysis of RFR

Forest plots of the random-effects meta-analysis of RFR are presented in Fig 3 for the TEOAE-TEOAE test combination and in Fig 4 for the other test combinations. A summary of the meta-analysis results is presented in Table 2. Forest plots for the meta-analysis of well-baby protocols are provided in the Supporting Information, S2 Fig and S3 Fig.

TEOAE-TEOAE

Study	Failures	Total screened	RFR	95% CI
Aidan et al. 1999	124	1421	0.087	[0.073; 0.103]
Alanazi 2020	504	20171	0.025	[0.023; 0.027]
Arjmandi et al. 2012	6	1232	0.005	[0.002; 0.011]
Arora et al. 2022	44	1200	0.037	[0.027; 0.049]
Arslan et al. 2013	55	2229	0.025	[0.019; 0.032]
Azizi et al. 2016	94	3818	0.025	[0.020; 0.030]
Benito-Orejas et al. 2008	64	2454	0.026	[0.020; 0.033]
Bevilacqua et al. 2010	885	11466	0.077	[0.072; 0.082]
Busse et al. 2023	47	778	0.060	[0.045; 0.080]
Calevo et al. 2007 (a)	58	3238	0.018	[0.014; 0.023]
Calevo et al. 2007	462	31796	0.015	[0.013; 0.016]
Cavalcanti et al. 2012	290	3724	0.078	[0.069; 0.087]
Chapchap et al. 2001	38	4196	0.009	[0.006; 0.012]
Clarke et al. 2003	27	81	0.333	[0.232; 0.447]
De Capua et al. 2007	447	19700	0.023	[0.021; 0.025]
Diego et al. 2023	107	1553	0.069	[0.057; 0.083]
Eibenstein et al. 2014	219	4579	0.048	[0.042; 0.054]
Eibenstein et al. 2015	47	3120	0.015	[0.011; 0.020]
Erturk et al. 2010	6	500	0.012	[0.004; 0.026]
Escobar-Ipuz et al. 2019	204	9390	0.022	[0.019; 0.025]
Farahani et al. 2017	8	2784	0.003	[0.001; 0.006]
Ferlito et al. 2021	361	37562	0.010	[0.009; 0.011]
Fusetti et al. 2007	55	1400	0.039	[0.030; 0.051]
George et al. 2016	9	1834	0.005	[0.002; 0.009]
Gilbey et al. 2013	261	4958	0.053	[0.047; 0.059]
Guastini et al. 2010	122	8671	0.014	[0.012; 0.017]
Gül et al. 2013	385	2363	0.163	[0.148; 0.178]
Habib et al. 2005	300	11986	0.025	[0.022; 0.028]
Hazopoulus et al. 2007	22	450	0.049	[0.031; 0.073]
Jakubikova et al. 2003	59	3048	0.019	[0.015; 0.025]
Kayiran et al. 2009	16	8052	0.002	[0.001; 0.003]
Konukseven et al. 2010	48	1917	0.025	[0.019; 0.033]
Korres et al. 2006 (1)	535	22195	0.024	[0.022; 0.026]
Korres et al. 2006 (2)	534	25032	0.021	[0.020; 0.023]
Kosmidou et al. 2021	13	1491	0.009	[0.005; 0.015]
Lin et al. 2004	231	5938	0.039	[0.034; 0.044]
Lotfi et al. 2007	163	7718	0.021	[0.018; 0.025]
Magnani et al. 2015	91	10359	0.009	[0.007; 0.011]
Molini et al. 2016	1006	18796	0.054	[0.050; 0.057]
Molteni 2006	209	10454	0.020	[0.017; 0.023]
Pastorino et al. 2005	175	19290	0.009	[0.008; 0.011]
Pedersen et al. 2008	98	1627	0.060	[0.049; 0.073]
Prpic et al. 2007	179	11746	0.015	[0.013; 0.018]
Pyarali et al. 2023	10	267	0.037	[0.018; 0.068]
Satish et al. 2019	408	26487	0.015	[0.014; 0.017]
Sennaroglu et al. 2011	150	1840	0.082	[0.069; 0.095]
Sequi Canet et al. 2016	182	14015	0.013	[0.011; 0.015]
Sheng et al. 2021	39	1340	0.029	[0.021; 0.040]
Tasci et al. 2010	322	15323	0.021	[0.019; 0.023]
Tatli et al. 2007	11	711	0.015	[0.008; 0.028]
Unlu et al. 2015	126	2933	0.043	[0.036; 0.051]
Vaid et al. 2009	35	1238	0.028	[0.020; 0.039]
Vos et al. 2014	12333	245219	0.050	[0.049; 0.051]
Welzl-Müller et al. 1997	120	2338	0.051	[0.043; 0.061]
Yorulmaz et al. 2017	1843	13693	0.135	[0.129; 0.140]
Random effects model	**24187**	**671721**	**0.027**	**[0.022; 0.032]**



Heterogeneity: I^2 = 39.4%, τ^2 = 0.564

Fig 3. Random-effects meta-analysis of refer rate (RFR) for the 55 TEOAE-TEOAE study protocols.

The table shows the number of newborns who did not pass the first and second test ("Failures"), the total number of screened newborns ("Total screened"), and the RFR with 95% confidence interval (95% CI) for each study. The summary estimate, including the 95% CI is shown as a gray diamond. The vertical solid line indicates the 4% threshold quality criteria defined in the Pediatrics Directive for the RFR. TEOAE = transitory evoked otoacoustic emission.

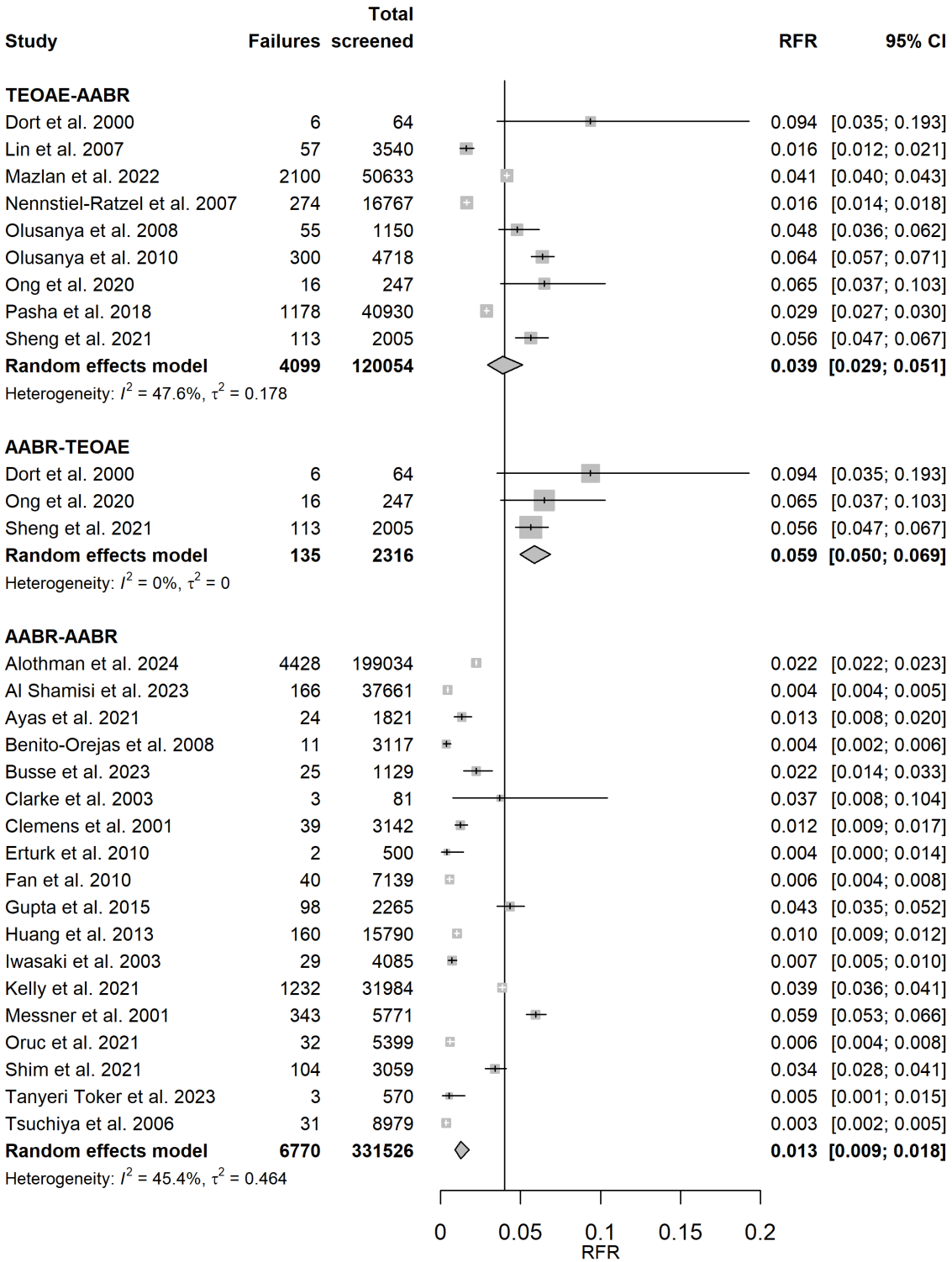


Fig 4. Random-effects meta-analysis of refer rate (RFR) for TEOAE-AABR, AABR-TEOAE, and AABR-AABR study protocols. The table shows the number of newborns who did not pass the first and second test ("Failures"), the total number of screened newborns ("Total screened"), and the RFR with 95% confidence interval (95% CI) for each study. The summary estimate per test combination, including the 95% CI, is shown as a gray diamond. The vertical solid line indicates the 4% threshold quality criteria defined in the Pediatrics

Directive for the RFR. AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission.

Table 2: Results of the random effects meta-analysis of the refer rate.

Test combination	**RFR (95% CI)**	**I^2**	**tau^2**	**Q**	**df**	**p**
TEOAE-TEOAE	2.7% (2.2, 3.2%)	39.4%	0.564	89.14	54	**0.002**
TEOAE-AABR	3.9% (2.9, 5.1%)	47.6%	0.178	15.27	8	0.054
AABR-TEOAE	5.9% (5.0, 6.9%)	0%	0	1.76	2	0.416
AABR-AABR	1.3% (0.9, 1.8%)	45.4%	0.464	31.14	17	**0.019**

RFR = refer rate, CI = confidence interval, AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission. The p-value shown in the right column relates to the Cochrane's Q statistic (Q) for heterogeneity. Statistically significant p values ($p<0.05$) are printed in bold.

Strategies that do not involve a change in the screening test method showed the lowest RFR. The upper limits of their 95% confidence intervals are below the recommended quality threshold of 4%. When the screening test method is changed between stage 1 and stage 2, the corresponding 95% confidence intervals cover or exceed the 4% threshold.

Studies of both screening combinations TEOAE-TEOAE and AABR-AABR show a moderate degree of heterogeneity as quantified by I^2 (39.4% and 45.4%, respectively). Excluding the TEOAE-TEOAE study protocol with a remarkably high RFR of 33.3% [26] and the two other studies with RFR > 10% [53,80], as shown in S4 Fig, reduces the summary estimate for RFR from 2.7% (2.2, 3.2%) to 2.4% (2.0, 2.8%). The results for RFR for well babies (S2 Fig and S3 Fig) are comparable to those of all studies: AABR-AABR and TEOAE-TEOAE show the lowest RFR (less than 4%). Changing the test method results in a higher RFR.

The results of the meta-regression of the failure rate of the first test and the RFR are presented in S1 Table and S5 Fig. The meta-regression regresses the RFR on the positive rate of the first stage (Eq.2) and allows inference on the conditional positivity rate in stage 2.

Meta-analysis of the pooled studies on the two-stage screening process to infer test-specific sensitivities and specificities

Estimates for the sensitivity and specificity of the two screening methods were calculated using our final model for the two-stage screening process and the methodology of Joseph et al. [18], (Table 3). It is assumed that the sensitivities and specificities for each diagnostic test are the same in all studies and the prevalence is 1.0 per 1000 children according to prior knowledge.

Table 3: Results for sensitivity and specificity across all 85 study protocols (middle column) and for the 38 study protocols that included only well babies (right column).

	All study protocols (n=85)	**Well-baby study protocols (n=38)**
Parameter	**Estimate (95% CrI)**	**Estimate (95% CrI)**
Sensitivity AABR	99.9% (99.6, 100.0%)	99.7% (98.7, 100.0%)
Sensitivity TEOAE	100.0% (99.9, 100.0%)	99.8% (99.4, 100.0%)
Specificity AABR	96.9% (96.8, 97.0%)	90.8% (90.5, 91.0%)
Specificity TEOAE	91.1% (91.0, 91.2%)	89.8% (89.7, 89.9%)

CrI = credibility interval, AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission.

Sensitivities of 99.9% (CrI: 99.6%, 100.0%) and 100.0% (CrI: 99.9%, 100.0%) were found for AABR and TEOAE, respectively. The corresponding specificities were 96.9% (CrI: 96.8%, 97.0%) and 91.1% (CrI: 91.0%, 91.2%). Both methods have very similar sensitivity. In contrast to this, the specificity of AABR is statistically significantly higher than that of TEOAE. Table 3 also shows the results of the subgroup analysis of 38 study protocols that included only well babies (right column). Compared to all studies (middle column), the sensitivities remained very similar and the specificities were slightly lower.

Bias assessment

Fig 5 shows the results of the bias assessment. While there was a low risk of bias in the use of the tests (first test: 98.8%, second test: 100%), about half of the studies had a high risk of bias in patient selection and in flow and timing (49.4% and 52.9%, respectively). In terms of applicability concerns, about 3 out of 4 studies had high concerns for patient selection (72.9%). Similar to the risk of bias, both tests showed only low applicability concerns (first test: 97.6%, second test: 100%). The study-level bias assessment for all domains can be found in S2 Table.

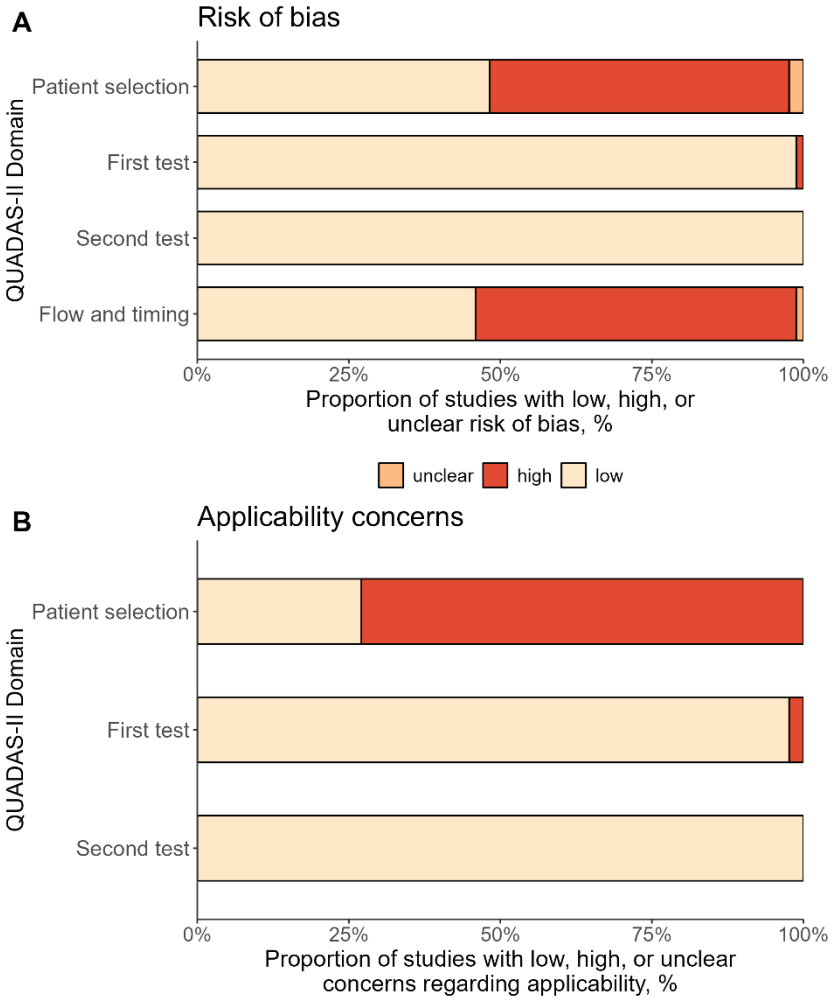


Fig 5. Bias assessment using QUADAS-II tool. The figure shows the percentage of the included 85 study protocols with low, high, or unclear risk of bias (A) and concerns regarding applicability (B).

The GRADE Summary of Findings table is included in the S1 File. We rated all evidence generated in this meta-analysis as moderate, which is the second highest GRADE evidence category. Specifically, the evidence for estimating sensitivities and specificities was lowered to moderate because of the assumption of equal sensitivities and specificities across studies. The level of evidence for the RFR from the random-effects meta-analysis was lowered to moderate because of the moderate heterogeneity found, and for test combinations with test method change between the two stages because of the small number of pooled studies.

Discussion

A meta-analysis was conducted to identify the optimal screening algorithm for a two-stage NHS using combinations of TEOAE or AABR tests. The study analyzed 85 study protocols with over 1,12 million newborns meeting the inclusion criteria. The results showed that the refer rate (RFR) was lower when there was no change in the screening method used. The aggregated RFR was 1.3% for the AABR-AABR test combination and 2.7% for TEOAE-TEOAE. The studies demonstrated high sensitivity and specificity for both the AABR and TEOAE tests.

The following discussion focuses on newborn hearing screening in the German setting. However, we believe that the discussion will be of broader interest to other countries following similar screening strategies, as the hearing screening process should always be evaluated and adjusted as necessary to achieve the best possible quality.

The RFR is an important quality parameter in the NHS, as screenings with a refer result must be followed up by a pediatric audiologist. These specialists are scarce in Germany (as well as in other countries), which can lead to long waiting times for families with children who have not passed the screening tests. In addition, false-positive results can cause unnecessary anxiety for parents [101]. A low RFR in the NHS can be achieved primarily through a multi-step screening algorithm. Therefore, when the NHS was introduced in 2009, the German Pediatrics Directive required an AABR control if the first hearing test was not passed and an RFR of less than 4% at discharge, in line with national and international quality targets at the time [14,102]. The UK even requested lower RFRs (acceptable: 3%, achievable: 2%) [17].

The German NHS evaluation data for the years 2011/2012 and the follow-up evaluation data for 2017/2018 both indicate that a second hearing test following an initial screening test with a “fail” result considerably reduces the RFR, as this second test was passed in over 80% [8,15]. However, it is worth noting, that this second measurement was performed with a TEOAE in more than half of the tests, which is contrary to the German Pediatrics Directive. In follow-up evaluation interviews this was explained by the longer measurement duration and the increased susceptibility to interference of the AABR. The follow-up evaluation analysis revealed a considerably higher rate of refer results in the second test when a different test method was used than in the first test [8,15]. This is in line with the finding of this meta-analysis that the highest RFR were observed when the test method was changed (TEOAE-AABR or AABR-TEOAE).

In the German follow-up evaluation, the TEOAE-TEOAE algorithm had a lower RFR (9.62%) than AABR-AABR (13.98%) [15]. In contrast, in this meta-analysis the AABR-AABR test combination had the lowest RFR, even lower than that of the TEOAE-TEOAE algorithm, while also having the highest specificity. This was also observed in the subgroup analysis on “well babies” and may be attributed to the higher lost-to-follow-up rate in the studies reporting TEOAE-TEOAE results (see Fig 2). In addition, studies utilizing this test combination exhibited a significantly higher heterogeneity (see Table 2, Q statistic) and were often based on routine clinical data, whereas data for the AABR-AABR test sequence were mainly derived from clinical studies. These factors may have affected the estimation of sensitivity and specificity. Nevertheless, the TEOAE-TEOAE test combination also had a low RFR while maintaining very good specificity and sensitivity.

One limitation of our study is the assumption of homogeneity of sensitivities and specificities across all studies, which ignores the heterogeneity caused by differently qualified personnel and in different settings (i.e. quiet vs. noisy). As these factors influencing the screening result are only described in detail in a few reports, they could not be considered in the meta-analysis. Similarly, the reports often lack information on whether data from children with risk factors for hearing impairment were included. However, they usually provide information on whether children from the NICU were included. In the subgroup analysis including exclusively study protocols with newborns without risk factors (“well babies”), there were only minor differences in the sensitivity and specificity of the different methods. However, the specificity of AABR

decreased by about 6% when including only well baby studies. As in the meta-analysis of all studies, the AABR-AABR and TEOAE-TEOAE algorithms showed the lowest RFR (below 4%) also in the subgroup of well babies, as well as higher RFR after changing the test method.

To our knowledge, this is the first systematic review on quality measures of tests used in a two-stage NHS. Strengths of the study include the large number of newborns included and the high number of study reports. While we analyzed the refer rate using a standard frequentist random-effects meta-analysis, our proposed method of estimating sensitivity and specificity using Bayesian methods may provide an interesting way for analysis in future studies. We found that the sensitivity of the tests increased and the specificity decreased as the prior knowledge of the prevalence of hearing impairment increased. The sensitivity of the AABR appeared to be more sensitive to changes in the prior belief about prevalence, whereas the sensitivity of the TEOAE appeared to be more stable. The specificities of AABR and TEOAE behaved similarly to changes in prior beliefs about the precision of the assumed prevalence of hearing impairment.

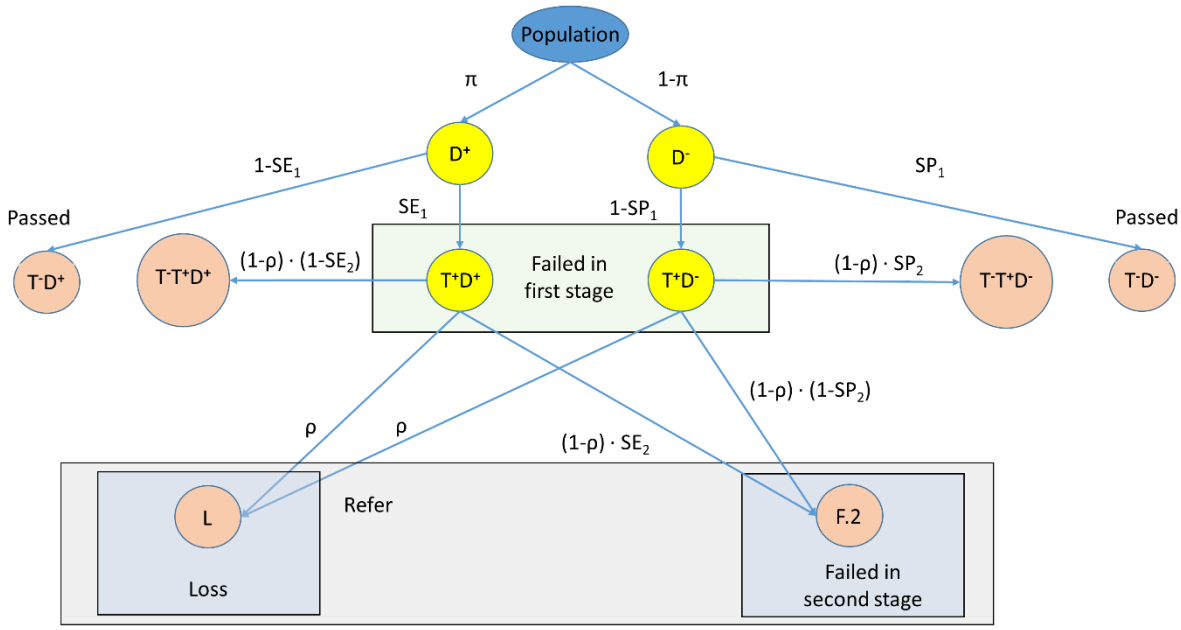
In its most recent 2019 position paper, the Joint Committee on Infant Hearing recommends performing at least two screening attempts with the same method or an AABR after TEOAE before discharge of a well baby. TEOAE testing after an initial AABR with a refer result is also acceptable for well babies, as the lost-to-follow-up rate for outpatient follow-up is very high [6,103]. The UK screening program guidelines recommend performing a further TEOAE test with an interval of at least 5 hours if the first TEOAE test is not passed for well babies [17].

The results of the meta-analysis and the data analysis of the German follow-up evaluation should provide evidence for adjusting the German Pediatrics Directive regarding the method of the second test to improve the RFR and align with international recommendations. Staff compliance with performing a second test before discharge is expected to improve if TEOAE tests are allowed, as TEOAE tests are faster and easier to perform than an AABR measurement. In contrast, changing the method of the second hearing test after the initial test was not passed results in a higher RFR and does not appear to have any benefits. However, in children with risk factors for perinatal hearing impairment, both hearing tests should always be performed with an AABR test, as specified in the German Pediatrics Directive [14] and international guidelines.

Acknowledgments

We would like to thank Anja Friedrichs for the full-text retrieval of non-open access studies.

Supporting Information



S1 Fig. Model for two-stage newborn hearing screening.
π = prevalence of hearing impairment, D+ (D-) = newborns with (without) hearing impairment, T- (T+) =newborns with "pass" ("fail") test results, SE_1 (SE_2) = sensitivity of the first (second) test, SP_1 (SP_2) = specificity of the first (second) test, ρ = loss rate after the first test stage, L = number of newborns lost after the first test result was "fail", F.2=number of newborns whose second test result was "fail".

S1 Checklist. PRISMA checklist.

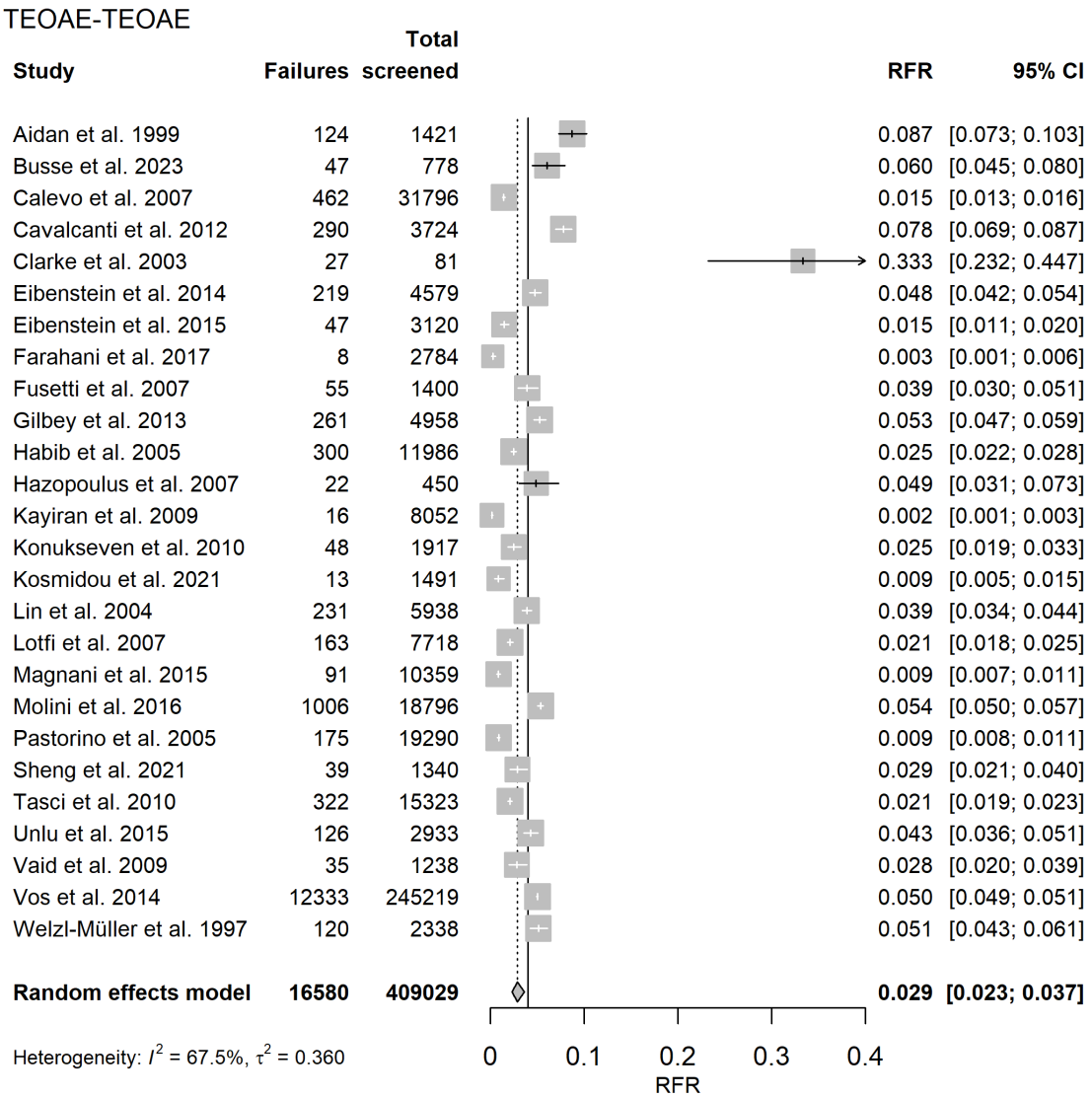
Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Considered in abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction (paragraphs 1-4)
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction (paragraph 4)
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Materials and methods → Eligibility criteria
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Materials and methods → Search method
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits	Materials and methods →

Section and Topic	Item #	Checklist item	Location where item is reported
		used.	Search method
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Materials and methods → Study records
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Materials and methods → Data extraction and items
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Materials and methods → Data extraction and items
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Materials and methods → Data extraction and items, legend of Table 1
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Materials and methods → Risk of bias in individual studies and publication bias
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Materials and methods → Data extraction and items
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Materials and methods → Statistical analysis
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Materials and methods → Data extraction and items
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Materials and methods → Statistical analysis
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Materials and methods → Statistical analysis
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Materials and methods → Subgroup analysis
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Materials and methods → Subgroup analysis

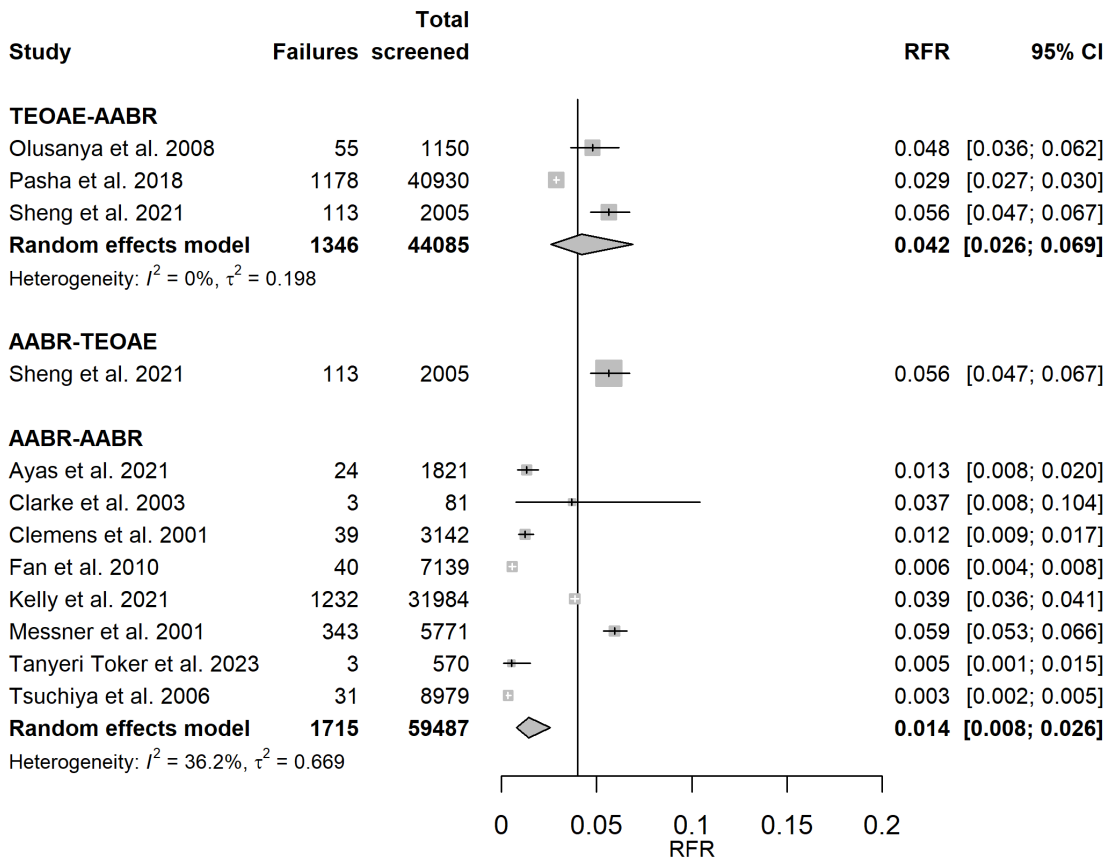
Section and Topic	Item #	Checklist item	Location where item is reported
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Materials and methods → Risk of bias in individual studies and publication bias
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Materials and methods → Confidence in cumulative evidence
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Fig 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Legend of Fig 1
Study characteristics	17	Cite each included study and present its characteristics.	Table 1
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	S2 Table
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Fig 2
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Fig 3
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	See Figs 3 and 4 and Tables 2 and 3
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	S4 Fig
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	S4 Fig
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	NA, since no reporting bias occurred.
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	S1 File
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion (paragraphs 2-4)
	23b	Discuss any limitations of the evidence included in the review.	Discussion (paragraph 5)
	23c	Discuss any limitations of the review processes used.	Discussion (paragraph 5)
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion (last paragraph)
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Materials and methods → Study protocol

Section and Topic	Item #	Checklist item	Location where item is reported
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Materials and methods → Study protocol
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Materials and methods → Study protocol
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	See Additional information
Competing interests	26	Declare any competing interests of review authors.	See Additional information
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Materials and Methods → Statistical analysis (last sentence)

NA = not applicable.



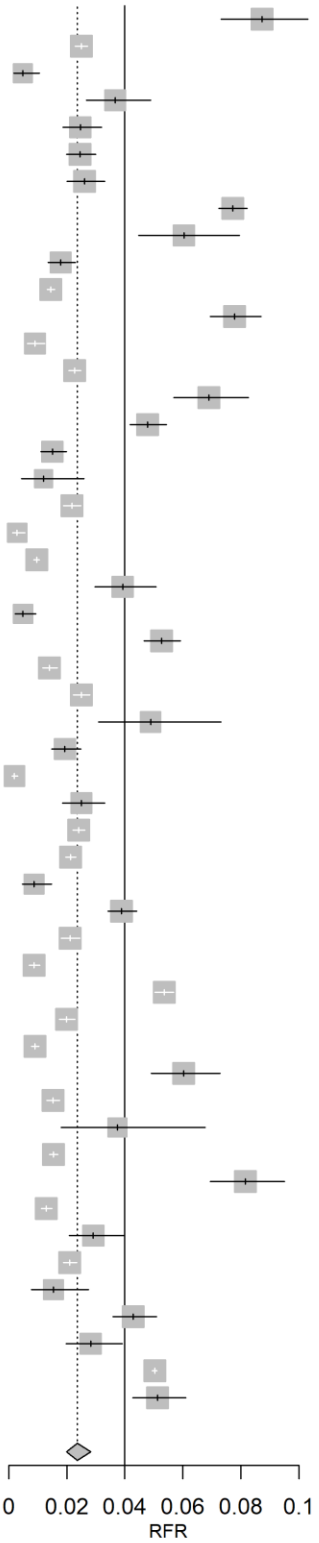
S2 Fig. Random-effects meta-analysis of the refer rate (RFR) for the 26 TEOAE-TEOAE well-baby study protocols. Shown are the number of newborns who failed both the first and second test ("Failures"), the number of screened newborns ("Total screened"), and the RFR with the 95% confidence interval (95% CI) for each study. The summary estimate including the 95% CI is shown as a gray diamond. The vertical solid line indicates the 4% threshold quality criteria defined in the Pediatrics Directive for the RFR. TEOAE = transitory evoked otoacoustic emission.



S3 Fig. Random-effects meta-analysis of the refer rate (RFR) for TEOAE-AABR, AABR-TEOAE and AABR-AABR well-baby study protocols. Shown are the number of newborns who did not pass first and second test ("Failures"), the number of screened newborns ("Total screened"), and the RFR with the 95% confidence interval (95% CI) for each study. The summary estimate per test combination including the 95% CI is shown as a gray diamond. The vertical solid line indicates the 4% threshold quality criteria defined in the Pediatrics Directive for the RFR. AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission.

TEOAE-TEOAE

Study	Failures	Total screened	RFR	95% CI
Aidan et al. 1999	124	1421	0.087	[0.073; 0.103]
Alanazi 2020	504	20171	0.025	[0.023; 0.027]
Arjmandi et al. 2012	6	1232	0.005	[0.002; 0.011]
Arora et al. 2022	44	1200	0.037	[0.027; 0.049]
Arslan et al. 2013	55	2229	0.025	[0.019; 0.032]
Azizi et al. 2016	94	3818	0.025	[0.020; 0.030]
Benito-Orejas et al. 2008	64	2454	0.026	[0.020; 0.033]
Bevilacqua et al. 2010	885	11466	0.077	[0.072; 0.082]
Busse et al. 2023	47	778	0.060	[0.045; 0.080]
Calevo et al. 2007 (a)	58	3238	0.018	[0.014; 0.023]
Calevo et al. 2007	462	31796	0.015	[0.013; 0.016]
Cavalcanti et al. 2012	290	3724	0.078	[0.069; 0.087]
Chapchap et al. 2001	38	4196	0.009	[0.006; 0.012]
De Capua et al. 2007	447	19700	0.023	[0.021; 0.025]
Diego et al. 2023	107	1553	0.069	[0.057; 0.083]
Eibenstein et al. 2014	219	4579	0.048	[0.042; 0.054]
Eibenstein et al. 2015	47	3120	0.015	[0.011; 0.020]
Erturk et al. 2010	6	500	0.012	[0.004; 0.026]
Escobar-Ipuz et al. 2019	204	9390	0.022	[0.019; 0.025]
Farahani et al. 2017	8	2784	0.003	[0.001; 0.006]
Ferlito et al. 2021	361	37562	0.010	[0.009; 0.011]
Fusetti et al. 2007	55	1400	0.039	[0.030; 0.051]
George et al. 2016	9	1834	0.005	[0.002; 0.009]
Gilbey et al. 2013	261	4958	0.053	[0.047; 0.059]
Guastini et al. 2010	122	8671	0.014	[0.012; 0.017]
Habib et al. 2005	300	11986	0.025	[0.022; 0.028]
Hazopoulus et al. 2007	22	450	0.049	[0.031; 0.073]
Jakubikova et al. 2003	59	3048	0.019	[0.015; 0.025]
Kayiran et al. 2009	16	8052	0.002	[0.001; 0.003]
Konukseven et al. 2010	48	1917	0.025	[0.019; 0.033]
Korres et al. 2006 (1)	535	22195	0.024	[0.022; 0.026]
Korres et al. 2006 (2)	534	25032	0.021	[0.020; 0.023]
Kosmidou et al. 2021	13	1491	0.009	[0.005; 0.015]
Lin et al. 2004	231	5938	0.039	[0.034; 0.044]
Lotfi et al. 2007	163	7718	0.021	[0.018; 0.025]
Magnani et al. 2015	91	10359	0.009	[0.007; 0.011]
Molini et al. 2016	1006	18796	0.054	[0.050; 0.057]
Molteni 2006	209	10454	0.020	[0.017; 0.023]
Pastorino et al. 2005	175	19290	0.009	[0.008; 0.011]
Pedersen et al. 2008	98	1627	0.060	[0.049; 0.073]
Prpic et al. 2007	179	11746	0.015	[0.013; 0.018]
Pyarali et al. 2023	10	267	0.037	[0.018; 0.068]
Satish et al. 2019	408	26487	0.015	[0.014; 0.017]
Sennaroglu et al. 2011	150	1840	0.082	[0.069; 0.095]
Sequi Canet et al. 2016	182	14015	0.013	[0.011; 0.015]
Sheng et al. 2021	39	1340	0.029	[0.021; 0.040]
Tasci et al. 2010	322	15323	0.021	[0.019; 0.023]
Tatli et al. 2007	11	711	0.015	[0.008; 0.028]
Unlu et al. 2015	126	2933	0.043	[0.036; 0.051]
Vaid et al. 2009	35	1238	0.028	[0.020; 0.039]
Vos et al. 2014	12333	245219	0.050	[0.049; 0.051]
Welzl-Müller et al. 1997	120	2338	0.051	[0.043; 0.061]
Random effects model	**21932**	**655584**	**0.024**	**[0.020; 0.028]**

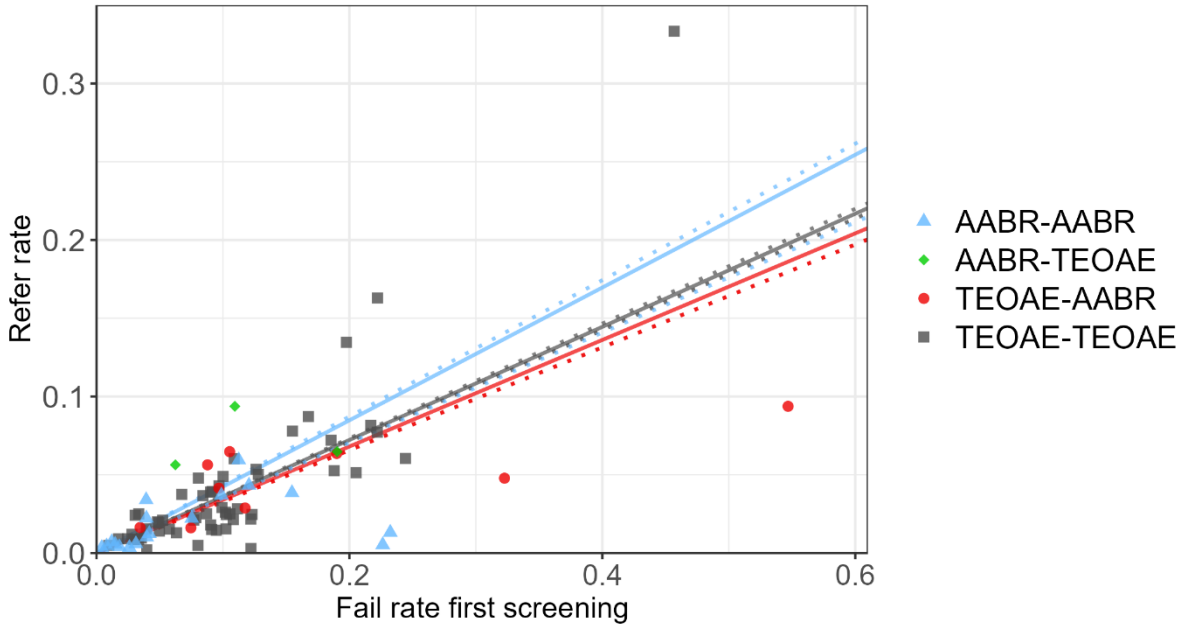


Heterogeneity: I^2 = 37.3%, τ^2 = 0.419

S4 Fig. Random-effects meta-analysis of refer rate (RFR) for 52 TEOAE-TEOAE study protocols without outliers. Excluding the studies by Clarke et al. 2003, Gül et al. 2013 and Yorulmaz et al. 2017. Shown are the number of newborns who did not pass first and second test ("Failures"), the number of screened newborns ("Total screened"), and the RFR with 95% confidence interval (95% CI) for each study. The summary estimate including the 95% CI is shown as a gray diamond. The vertical solid line indicates the 4% threshold quality criteria defined in the Pediatrics Directive for the RFR. TEOAE = transitory evoked otoacoustic emission.

Test combination	Mean slope (95% CrI)
TEOAE-TEOAE	0.361 (0.356, 0.367)
TEOAE-AABR	0.340 (0.329, 0.352)
AABR-AABR	0.424 (0.412, 0.436)

S1 Table: Results of the meta-regression.
SD = standard deviation, CrI = credibility interval, AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission.



S5 Fig. Refer rates in dependence of the failure rate of the first test for the test combinations AABR-AABR (n=18, blue triangles), AABR-TEOAE (n=3, green diamonds), TEOAE-AABR (n=9, red dots) and TEOAE-TEOAE (n=55, gray squares). AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission. Meta-regression lines including credibility intervals are plotted with data from S1 Table. No meta-regression line is shown for the AABR-TEOAE test combination with only three studies.

	Risk of bias				Applicability concerns		
Study	**Patient selection**	**First test**	**Second test**	**Flow and timing**	**Patient selection**	**First test**	**Second test**
TEOAE-TEOAE:							
Aidan et al. 1999	low	low	low	high	high	low	low
Alanazi 2020	low	low	low	high	high	low	low
Arjmandi et al. 2012	high	low	low	high	high	Low	low
Arora et al. 2022	high	low	low	high	high	low	low
Arslan et al. 2013	high	low	low	low	high	low	low
Azizi et al. 2016	high	low	low	high	high	low	low
Benito-Orejas et al. 2008	high	low	low	high	high	low	low
Bevilacqua et al. 2021	low	low	low	high	high	low	low
Busse et al. 2023	low	low	low	high	high	low	low
Calevo et al. 2007 (a)	unclear	low	low	low	low	low	low
Calevo et al. 2007	low	low	low	low	low	low	low
Calvacanti et al. 2012	high	low	low	high	high	low	low
Chapchap et al. 2001	high	low	low	high	high	low	low
Clarke et al. 2003	high	low	low	low	low	low	low
De Capua et al. 2007	high	low	low	high	high	low	low
Diego et al. 2023	high	low	low	high	high	low	low
Eibenstein et al. 2014	low	low	low	high	high	low	low
Eibenstein et al 2015	low	low	low	high	low	low	low
Erturk et al. 2010	unclear	low	low	high	high	low	low
Escobar-Ipuz et al. 2019	high	low	low	low	high	low	low
Farahani et al. 2017	low	low	low	low	low	low	low
Ferlito et al. 2021	high	low	low	high	high	low	low
George et al. 2016	high	low	low	high	high	low	low
Fusetti et al. 2007	low	low	low	high	high	low	low
Gilbey et al. 2013	low	low	low	low	low	low	low
Guastini et al. 2010	high	low	low	low	high	low	low
Gül et al. 2013	high	low	low	high	high	low	low
Habib et al. 2005	low	low	low	low	low	low	low
Hatzopoulus et al. 2007	low	low	low	high	high	low	low
Jakubikova et al. 2003	high	low	low	high	high	low	low
Kayiran et al. 2009	high	low	low	low	high	low	low
Konukseven et al. 2010	low	low	low	high	high	low	low
Korres et al. 2006 (1)	high	high	low	high	high	high	low
Korres et al. 2006 (2)	low	low	low	low	low	low	low
Kosmidou et al. 2021	low	low	low	low	low	low	low
Lin et al. 2004	high	low	low	high	high	low	low
Lofti et al. 2007	low	low	low	high	low	low	low
Magnani et al. 2015	low	low	low	low	low	low	low
Molini et al. 2016	low	low	low	high	high	low	low
Molteni 2006	high	low	low	high	high	low	low
Pastorino et al. 2005	low	low	low	high	high	high	low
Pedersen et al. 2008	high	low	low	high	high	low	low
Prpic et al. 2007	high	low	low	low	high	low	low
Pyarali et al. 2023	high	low	low	low	high	low	low
Satish et al. 2019	high	low	low	low	high	low	low
Sennaroglu et al. 2011	high	low	low	unclear	high	low	low
Sequi Canet et al. 2016	low	low	low	high	high	low	low
Sheng et al. 2021	low	low	low	high	high	low	low
Tasci et al. 2010	low	low	low	low	high	low	low
Tatli et al. 2007	high	low	low	high	high	low	low

	Risk of bias				Applicability concerns		
Study	**Patient selection**	**First test**	**Second test**	**Flow and timing**	**Patient selection**	**First test**	**Second test**
Unlu et al. 2015	low	low	low	low	low	low	low
Vaid et al. 2009	low	low	low	high	high	low	low
Vos et al. 2014	low	low	low	high	high	low	low
Welzl-Müller et al. 1997	low	low	low	high	high	low	low
Yorulmaz et al. 2017	low	low	low	high	high	low	low
TEOAE-AABR:							
Dort et al. 2000	high	low	low	low	high	low	low
Lin et al. 2007	low	low	low	low	low	low	low
Mazlan et al. 2022	low	low	low	high	high	low	low
Nennstiel-Ratzel et al. 2007	high	low	low	high	high	low	low
Olusanya et al. 2008	low	low	low	low	low	low	low
Olusanya et al. 2010	high	low	low	high	high	low	low
Ong et al. 2020	high	low	low	low	high	low	low
Pasha et al. 2018	low	low	low	high	high	low	low
Sheng et al. 2021	low	low	low	low	low	low	low
AABR-TEOAE:							
Dort et al. 2000	high	low	low	low	high	low	low
Ong et al. 2020	low	low	low	low	high	low	low
Sheng et al. 2021	low	low	low	low	low	low	low
AABR-AABR:							
Alothman et al. 2024	high	low	low	low	high	low	low
Al Shamisi et al. 2023	high	low	low	high	high	low	low
Ayas et al. 2021	low	low	low	low	low	low	low
Benito-Orejas et al. 2008	low	low	low	high	high	low	low
Busse et al. 2023	high	low	low	low	high	low	low
Clarke et al. 2003	high	low	low	low	low	low	low
Clemens et al. 2001	low	low	low	low	low	low	low
Erturk et al. 2010	high	low	low	high	high	low	low
Fan et al. 2010	high	low	low	low	high	low	low
Gupta et al. 2015	high	low	low	low	high	low	low
Huang et al. 2013	high	low	low	low	high	low	low
Iwasaki et al. 2003	low	low	low	low	low	low	low
Kelly et al. 2021	low	low	low	low	low	low	low
Messner et al. 2001	low	low	low	high	high	low	low
Oruc et al. 2021	high	low	low	low	high	low	low
Shim et al. 2021	high	low	low	low	high	low	low
Tanyeri Toker et al. 2023	low	low	low	low	low	low	low
Tsuchiya et al. 2006	high	low	low	high	low	low	low

S2 Table: Study-level assessment of bias for all 85 study protocols.
AABR = automated auditory brainstem response, TEOAE = transitory evoked otoacoustic emission.

Population: (Well) babies undergoing a two-stage hearing screening
Setting: Newborns from studies from all countries (high, middle and low-income countries) with the following criteria
(1) Initial screening as an inpatient in the maternity clinic
(2) Second test up to a maximum of one month later
(3) No use of distortion product otoacoustic emissions (DPOAE)
(4) Not exclusively NICU (neonatal intensive care unit).
Includes studies of well babies only and studies of well babies and NICU infants. Studies of exclusively NICU infants were not included in this review.
Intervention: Two-stage hearing screening using TEOAE, AABR, or combination of both
Comparison: not applicable

Test findings			**Certainty of the evidence (GRADE)**	**Comments**
AABR	Pooled sensitivity (95% CrI)	99.9% (99.6, 100.0%)	Moderate[1]	Sensitivity of a test for a condition with low prevalence needs to be high, assumption of equal sensitivities across all studies
	Pooled specificity (95% CrI)	96.9% (96.8, 97.0%)	Moderate[2]	Assumption of equal specificities across all studies
TEOAE	Pooled sensitivity (95% CrI)	100.0% (99.9, 100.0%)	Moderate[1]	Sensitivity of a test for a condition with low prevalence needs to be high, assumption of equal sensitivities across all studies
	Pooled specificity (95% CrI)	91.1% (91.0, 91.2%)	Moderate[2]	Assumption of equal specificities across all studies
Screening algorithm	**Pooled refer rate (95% CI)**	**Number of newborns (studies)**		
AABR-AABR	1.3% (0.9, 1.8%)	331526 (18)	Moderate[3]	Moderate heterogeneity, but larger number of studies
TEOAE-TEOAE	2.7% (2.2, 3.2%)	671721 (55)	Moderate[3]	Moderate heterogeneity, but large number of studies
TEOAE-AABR	3.9% (2.9, 5.1%)	120054 (9)	Moderate[4]	Moderate heterogeneity and small number of studies
AABR-TEOAE	5.9% (5.0, 6.9%)	2316 (3)	Moderate[4]	No evidence of heterogeneity, but small number of studies

S1 File: Summary of findings table. CrI = Credibility interval, CI = confidence interval, GRADE = GRADE Working Group grades of evidence (high, moderate, low, very low).[1]Level of evidence was lowered by one level due to assumption of equal sensitivities in Bayesian analysis. [2]Level of evidence was lowered by one level due to assumption of equal specificities in Bayesian analysis. [3]Level of evidence was lowered by one level due to found moderate heterogeneity in random-effects meta-analysis. [4]Level of evidence was lowered by one level due to small number of pooled studies in random-effects meta-analysis.

References

1. Korver AMH, Konings S, Dekker FW, Beers M, Wever CC, Frijns JHM, et al. Newborn hearing screening vs later hearing screening and developmental outcomes in children with permanent childhood hearing impairment. JAMA. 2010;304: 1701-1708. doi: 10.1001/jama.2010.1501.

2. McCann DC, Worsfold S, Law CM, Mullee M, Petrou S, Stevenson J, et al. Reading and communication skills after universal newborn screening for permanent childhood hearing impairment. Arch Dis Child. 2009;94: 293-297. doi: 10.1136/adc.2008.151217.

3. Pimperton H, Blythe H, Kreppner J, Mahon M, Peacock JL, Stevenson J, et al. The impact of universal newborn hearing screening on long-term literacy outcomes: a prospective cohort study. Arch Dis Child. 2016;101: 9-15. doi: 10.1136/archdischild-2014-307516.

4. Wolff R, Hommerich J, Riemsma R, Antes G, Lange S, Kleijnen J. Hearing screening in newborns: systematic review of accuracy, effectiveness, and effects of interventions after screening. Arch Dis Child. 2010;95: 130-135. doi: 10.1136/adc.2008.151092.

5. Yoshinaga-Itano C, Manchaiah V, Hunnicutt C. Outcomes of Universal Newborn Screening Programs: Systematic Review. J Clin Med. 2021;10: 2784. doi: 10.3390/jcm10132784. doi: 10.3390/jcm10132784.

6. Joint committee on infant hearing. Year 2019 Position Statement: Principles and Guidelines for Early Hearing Detection and Intervention Programs. Journal of Early Hearing Detection and Intervention. 2019;4: 1-44. doi: 10.15142/fptk-b748.

7. Grandori F. European Consensus Statement on Neonatal Hearing Screening. J Laryngol Otol. 1998;112: 1219. doi: 10.1017/s002221510014294x.

8. Nennstiel U, Brockow I, Söhl K, Zirngibl A, am Zehnhoff-Dinnesen A, Matulat P, et al. [Endbericht zur Evaluation des Neugeborenen-Hörscreenings 2011/2012]. 2017. [cited Apr 3, 2024]. Available from: https://www.g-ba.de/downloads/40-268-4395/2017-05-18_Kinder-RL_Annahme_Endbericht_NHS-Bericht.pdf.

9. Grosse SD, Mason CA, Gaffney M, Thomson V, White KR. What Contribution Did Economic Evidence Make to the Adoption of Universal Newborn Hearing Screening Policies in the United States? Int J Neonatal Screen. 2018;4: 25. doi: 10.3390/ijns4030025.

10. Rohlfs A, Wiesner T, Drews H, Muller F, Breitfuss A, Schiller R, et al. Interdisciplinary approach to design, performance, and quality management in a multicenter newborn hearing screening project. Discussion of the results of newborn hearing screening in Hamburg (part II). Eur J Pediatr. 2010;169: 1453-1463. doi: 10.1007/s00431-010-1229-0.

11. Sharma R, Gu Y, Ching TYC, Marnane V, Parkinson B. Economic Evaluations of Childhood Hearing Loss Screening Programmes: A Systematic Review and Critique. Appl Health Econ Health Policy. 2019;17: 331-357. doi: 10.1007/s40258-018-00456-1.

12. Wilson JMG, Jungner G, World Health Organisation. Principles and practice of screening for disease. Geneva: World Health Organization; 1968.

13. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ. 2008;86: 317-319. doi: 10.2471/blt.07.050112.

14. Gemeinsamer Bundesausschuss. [Richtlinie des Gemeinsamen Bundesausschusses über die Früherkennung von Krankheiten bei Kindern (Kinder-Richtlinie)]. 2023. [cited Apr 3, 2024]. Available from: https://www.g-ba.de/downloads/62-492-3190/Kinder-RL_2023-05-12_iK-2023-07-13.pdf.

15. Nennstiel U, Brockow I, Hanauer M, Heißenhuber A, Marzi C, Söhl K, et al. [Endbericht zur Folge-Evaluation des Neugeborenen-Hörscreenings 2017/2018]. 2022. [cited Apr 3, 2024]. Available from: https://www.g-ba.de/downloads/40-268-9045/2022-11-17_Kinder-RL_Abnahme-Endbericht-Folge-Evaluation-NHS_Bericht.pdf.

16. Brockow I, Sohl K, Hanauer M, Heissenhuber A, Marzi C, Am Zehnhoff-Dinnesen A, et al. Newborn hearing screening in Germany-results of the 2011/2012 and 2017/2018 evaluations. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2023;66: 1259-1267. doi: 10.1007/s00103-023-03779-0.

17. Government Digital Service UK. Newborn hearing screening programme (NHSP) operational guidance. Guidance 6: Patient journey from screen to referral. 2022. [cited Apr 3, 2024]. Available from: https://www.gov.uk/government/publications/newborn-hearing-screening-programme-nhsp-operational-guidance/6-patient-journey-from-screen-to-referral.

18. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol. 1995;141: 263-272. doi: 10.1093/oxfordjournals.aje.a117428.

19. Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess. 2007;11: iii, ix-51. doi: 10.3310/hta11500.

20. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155: 529-536. doi: 10.7326/0003-4819-155-8-201110180-00009.

21. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336: 924-926. doi: 10.1136/bmj.39489.470347.AD.

22. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing. 2000;10: 325-337. doi: 10.1023/A:1008929526011.

23. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2023. [cited Apr 3, 2024]. Available from: https://www.R-project.org/.

24. Benito-Orejas JI, Ramirez B, Morais D, Almaraz A, Fernandez-Calvo JL. Comparison of two-step transient evoked otoacoustic emissions (TEOAE) and automated auditory brainstem response (AABR) for universal newborn hearing screening programs. Int J Pediatr Otorhinolaryngol. 2008;72: 1193-1201. doi: 10.1016/j.ijporl.2008.04.011.

25. Busse AML, Qirjazi B, Mackey AR, Kik J, Goedegebure A, Hoeve HLJ, et al. Implementation of Newborn Hearing Screening in Albania. Int J Neonatal Screen. 2023;9: 28. doi: 10.3390/ijns9020028. doi: 10.3390/ijns9020028.

26. Clarke P, Iqbal M, Mitchell S. A comparison of transient-evoked otoacoustic emissions and automated auditory brainstem responses for pre-discharge neonatal hearing screening. Int J Audiol. 2003;42: 443-447. doi: 10.3109/14992020309081514.

27. Dort JC, Tobolski C, Brown D. Screening strategies for neonatal hearing loss: which test is best? J Otolaryngol. 2000;29: 206-210.

28. Erturk BB, Genc GA, Ozkan S. Comparison of Hearing Screening Protocols for Universal Newborn Hearing Screening In Turkey. Journal of International Advanced Otology. 2010;6: 223-230.

29. Ong KMC, Rivera AS, Chan AL, Chiong CM. Determining concordance and cost impact of otoacoustic emission and automated auditory brainstem response in newborn hearing screening in a tertiary hospital. Int J Pediatr Otorhinolaryngol. 2020;128: 109704. doi: 10.1016/j.ijporl.2019.109704.

30. Sheng H, Zhou Q, Wang Q, Yu Y, Liu L, Liang M, et al. Comparison of Two-Step Transient Evoked Otoacoustic Emissions and One-Step Automated Auditory Brainstem Response for Universal Newborn Hearing Screening Programs in Remote Areas of China. Front Pediatr. 2021;9: 655625. doi: 10.3389/fped.2021.655625.

31. Aidan D, Avan P, Bonfils P. Auditory screening in neonates by means of transient evoked otoacoustic emissions: a report of 2,842 recordings. Ann Otol Rhinol Laryngol. 1999;108: 525-531. doi: 10.1177/000348949910800601.

32. Alanazi AA. Referral and Lost to System Rates of Two Newborn Hearing Screening Programs in Saudi Arabia. Int J Neonatal Screen. 2020;6: 50. doi: 10.3390/ijns6030050.

33. Arjmandi F, Fahangfar B, Mehrabi S, Toghiani A, Sohrabi H. Prevalence of deafness and hearing screening in newborns in Isfahan. J Res Med Sci. 2012;17: S233-S236.

34. Arora RD, Jati M, Nagarkar NM, Galhotra A, Agrawal S, Mehta R, et al. Experience, Challenges and Outcome of Implementing Universal New Born Hearing Screening in a Medical College Hospital Set Up. Indian J Otolaryngol Head Neck Surg. 2022;74: 3841-3846. doi: 10.1007/s12070-021-02633-6.

35. Arslan S, Isik AU, Imamoglu M, Topbas M, Aslan Y, Ural A. Universal newborn hearing screening; automated transient evoked otoacoustic emissions. B-ENT. 2013;9: 122-131.

36. Azizi A, Amirian F, Dargahi A, Beidaghi S, Mohammadi M. Evaluation of universal newborn hearing screening with TEOAE and ABR: A cross-sectional study with the literature review. International Journal of Tropical Medicine. 2016;11: 84-89.

37. Bevilacqua MC, Alvarenga KdF, Costa OA, Moret ALM. The universal newborn hearing screening in Brazil: from identification to intervention. Int J Pediatr Otorhinolaryngol. 2010;74: 510-515. doi: 10.1016/j.ijporl.2010.02.009.

38. Calevo MG, Mezzano P, Zullino E, Padovani P, Scopesi F, Serra G, et al. Neonatal hearing screening model: an Italian regional experience. J Matern Fetal Neonatal Med. 2007;20: 441-448. doi: 10.1080/14767050701398090.

39. Calevo MG, Mezzano P, Zullino E, Padovani P, Serra G, STERN Group. Ligurian experience on neonatal hearing screening: clinical and epidemiological aspects. Acta Paediatr. 2007;96: 1592-1599. doi: 10.1111/j.1651-2227.2007.00475.x.

40. Cavalcanti HG, Guerra RO. The role of maternal socioeconomic factors in the commitment to universal newborn hearing screening in the Northeastern region of Brazil. Int J Pediatr Otorhinolaryngol. 2012;76: 1661-1667. doi: 10.1016/j.ijporl.2012.07.041.

41. Chapchap MJ, Segre CM. Universal newborn hearing screening and transient evoked otoacoustic emission: new concepts in Brazil. Scand Audiol Suppl. 2001;2: 33-36. doi: 10.1080/010503901750166600.

42. De Capua B, Costantini D, Martufi C, Latini G, Gentile M, De Felice C. Universal neonatal hearing screening: the Siena (Italy) experience on 19,700 newborns. Early Hum Dev. 2007;83: 601-606. doi: 10.1016/j.earlhumdev.2007.01.001.

43. Diego Gimenes Lopes J, Disconzi Dallegrave C, Hellmann Delfino N, Lauxen R, Marcelino T, Eduardo Monteiro Zappelini C. Epidemiological Profile of Neonates in Hearing Screening at a Maternity of a Tertiary Hospital in the state of Santa Catarina, Brazil. Int Arch Otorhinolaryngol. 2023;27: e412-e422. doi: 10.1055/s-0043-1770918.

44. Eibenstein A, Varakliotis T, Saccoccio A, Cisternino S, Gamma R, Mattei A, et al. Newborn hearing screening with TEOAE: Our experience on 4759 infants. Otorinolaringologia. 2014;64: 27-32.

45. Eibenstein A, Saccoccio A, Di Rubbo V, Rosati N, Varakliotis T, Tizio A, et al. Newborn hearing screening program: our experience. Otorinolaringologia. 2015;65: 97-101.

46. Escobar-Ipuz FA, Soria-Bretones C, Garcia-Jimenez MA, Cueto EM, Torres Aranda AM, Sotos JM. Early detection of neonatal hearing loss by otoacoustic emissions and auditory brainstem response over 10 years of experience. Int J Pediatr Otorhinolaryngol. 2019;127: 109647. doi: 10.1016/j.ijporl.2019.109647.

47. Farahani F, Hamidi Nahrani M, Seifrabiei MA, Emadi M. The Effect of Mode of Delivery and Hospital Type on Newborn Hearing Screening Results Using Otoacoustic Emissions: Based on Screening Age. Indian J Otolaryngol Head Neck Surg. 2017;69: 1-5. doi: 10.1007/s12070-016-0967-3.

48. Ferlito S, Maniaci A, Cocuzza S, La Mantia I, Di Mauro P, Poli G, et al. Universal newborn hearing screening in the Italian Region of Sicily in 2018. Acta Otorhinolaryngol Ital. 2021;41: 356-363. doi: 10.14639/0392-100X-N1162.

49. George SA, Alawadhi A, Al Reefy H, Riskalla A. Newborn hearing screening. Bahrain Medical Bulletin. 2016;38: 148-150.

50. Fusetti M, Fioretti A, Ottaviano I, Simaskou M, Eibenstein A, Melpignano G. Newborn hearing screening with otoacoustic emissions: Our experience. Otorinolaringologia. 2007;57: 99-104.

51. Gilbey P, Kraus C, Ghanayim R, Sharabi-Nov A, Bretler S. Universal newborn hearing screening in Zefat, Israel: the first two years. Int J Pediatr Otorhinolaryngol. 2013;77: 97-100. doi: 10.1016/j.ijporl.2012.10.004.

52. Guastini L, Mora R, Dellepiane M, Santomauro V, Mora M, Rocca A, et al. Evaluation of an automated auditory brainstem response in a multi-stage infant hearing screening. Eur Arch Otorhinolaryngol. 2010;267: 1199-1205. doi: 10.1007/s00405-010-1209-z.

53. Gül A, Aliosmanoglu C, Cengul E. The results of newborn hearing screening in Diyarbakir Children's Hospital. Duzce Medical Journal. 2013;15: 4-6.

54. Habib HS, Abdelgaffar H. Neonatal hearing screening with transient evoked otoacoustic emissions in Western Saudi Arabia. Int J Pediatr Otorhinolaryngol. 2005;69: 839-842. doi: 10.1016/j.ijporl.2005.01.018.

55. Hatzopoulos S, Qirjazi B, Martini A. Neonatal hearing screening in Albania: results from an ongoing universal screening program. Int J Audiol. 2007;46: 176-182. doi: 10.1080/14992020601145310.

56. Jakubikova J, Kabatova Z, Zavodna M. Identification of hearing loss in newborns by transient otoacoustic emissions. Int J Pediatr Otorhinolaryngol. 2003;67: 15-18. doi: 10.1016/s0165-5876(02)00285-9.

57. Kayiran SM, Genc E, Erdil A, Gurakan B. Results of American Hospital newborn hearing screening program. Turk Arch Ped. 2009;44: 135-7.

58. Konukseven O, Genc A, Muderris T, Kayikci MK, Turkyilmaz D, Ozturk B, et al. Can automated auditory brainstem response be used as an initial stage screening test in newborn hearing screening programs? Journal of International Advanced Otology. 2010;6: 231-238.

59. Korres SG, Balatsouras DG, Nikolopoulos T, Korres GS, Ferekidis E. Making universal newborn hearing screening a success. Int J Pediatr Otorhinolaryngol. 2006;70: 241-246. doi: 10.1016/j.ijporl.2005.06.010.

60. Korres SG, Balatsouras DG, Gkoritsa E, Eliopoulos P, Rallis E, Ferekidis E. Success rate of newborn and follow-up screening of hearing using otoacoustic emissions. Int J Pediatr Otorhinolaryngol. 2006;70: 1039-1043. doi: 10.1016/j.ijporl.2005.10.018.

61. Kosmidou P, Tzifas S, Lygeros S, Danielides G, Nikolopoulos T, Dimitriou G, et al. Newborn Hearing Screening: Analysing the Effectiveness of Early Detection of Neonatal Hearing Loss in a Hospital in Greece. Cureus. 2021;13: e19807. doi: 10.7759/cureus.19807.

62. Lin C, Huang C, Lin C, Lin Y, Wu J. Community-based newborn hearing screening program in Taiwan. Int J Pediatr Otorhinolaryngol. 2004;68: 185-189. doi: 10.1016/j.ijporl.2003.10.007.

63. Lotfi Y, Movallali G. A universal newborn hearing screening in Iran. Iranian Rehabilitation Journal. 2007;5: 8-11.

64. Magnani C, Bacchi G, Borghini AM, Delmonte D, Fava G, Occasio AM, et al. Universal newborn hearing screening: the experience of the University Hospital of Parma. Acta Biomed. 2015;86: 273-277.

65. Molini E, Calzolaro L, Lapenna R, Ricci G. Universal newborn hearing screening in Umbria region, Italy. Int J Pediatr Otorhinolaryngol. 2016;82: 92-97. doi: 10.1016/j.ijporl.2016.01.007.

66. Molteni G. Early detection of newborn hearing impairment by transiently evoked otoacoustic emissions and auditory evoked potentials. Personal experience in 10,454 children. Otorinolaringologia. 2006;56: 93-6.

67. Pastorino G, Sergi P, Mastrangelo M, Ravazzani P, Tognola G, Parazzini M, et al. The Milan Project: a newborn hearing screening programme. Acta Paediatr. 2005;94: 458-463. doi: 10.1111/j.1651-2227.2005.tb01918.x.

68. Pedersen L, Møller TR, Wetke R, Ovesen T. [Neonatal hearing screening. A comparison of automatic auditory brainstem audiometry and otoacoustic emissions]. Ugeskr Laeger. 2008;170: 642-646.

69. Prpic I, Mahulja-Stamenkovic V, Bilic I, Haller H. Hearing loss assessed by universal newborn hearing screening--the new approach. Int J Pediatr Otorhinolaryngol. 2007;71: 1757-1761. doi: 10.1016/j.ijporl.2007.07.015.

70. Pyarali M, Akhtar S, Adeel M, Mallick SA, Uneeb SN, Aslam A. Neonatal Hearing Screening Programme And Challenges Faced By The Developing Country: A Tertiary Care Hospital Experience. J Pak Med Assoc. 2023;73: 1788-1793. doi: 10.47391/JPMA.6264.

71. Satish HS, Anil Kumar R, Viswanatha B. Screening of Newborn Hearing at a Tertiary Care Hospital in South India. Indian J Otolaryngol Head Neck Surg. 2019;71: 1383-1390. doi: 10.1007/s12070-018-1454-9.

72. Sennaroglu G, Akmese PP. Risk Factors for Hearing Loss and Results of Newborn Hearing Screening in Rural Area. Journal of International Advanced Otology. 2011;7: 343-350.

73. Sequi Canet JM, Sala Langa MJ, Collar Del Castillo JI. Results from ten years newborn hearing screening in a secondary hospital. An Pediatr (Barc). 2016;85: 189-196. doi: 10.1016/j.anpedi.2015.11.003.

74. Tasci Y, Muderris II, Erkaya S, Altinbas S, Yucel H, Haberal A. Newborn hearing screening programme outcomes in a research hospital from Turkey. Child Care Health Dev. 2010;36: 317-322. doi: 10.1111/j.1365-2214.2009.01029.x.

75. Tatli MM, Bulent Serbetcioglu M, Duman N, Kumral A, Kirkim G, Ogun B, et al. Feasibility of neonatal hearing screening program with two-stage transient otoacoustic emissions in Turkey. Pediatr Int. 2007;49: 161-166. doi: 10.1111/j.1442-200X.2007.02344.x.

76. Unlu I, Guclu E, Yaman H. When should automatic Auditory Brainstem Response test be used for newborn hearing screening? Auris Nasus Larynx. 2015;42: 199-202. doi: 10.1016/j.anl.2014.10.005.

77. Vaid N, Shanbhag J, Nikam R, Biswas A, Ears B. Neonatal hearing screening–The Indian experience. Cochlear implants international. 2009;10: 111-114. doi: 10.1179/cim.2009.10.Supplement-1.111.

78. Vos B, Lagasse R, Leveque A. Main outcomes of a newborn hearing screening program in Belgium over six years. Int J Pediatr Otorhinolaryngol. 2014;78: 1496-1502. doi: 10.1016/j.ijporl.2014.06.019.

79. Welzl-Muller K, Boheim K, Stephan K, Schlogel H, Stadlmann A, Nekahm D. Optimizing hearing screening by transient evoked otoacoustic emissions in newborn infants. HNO. 1997;45: 227-232. doi: 10.1007/s001060050109.

80. Yorulmaz A, Genç U, Yılmaz FH, Küçüksümbül S. Evaluation and Importance of Our Newborn Hearing Screening Results. Medical Bulletin of Haseki/Haseki Tip Bulteni. 2017;55.

81. Lin H, Shu M, Lee K, Lin H, Lin G. Reducing False Positives in Newborn Hearing Screening Program: how and Why. Otol Neurotol. 2007;28: 788-792. doi: 10.1097/mao.0b013e3180cab754.

82. Mazlan R, Raman K, Abdullah A. A 10-year retrospective analysis of newborn hearing screening in a tertiary hospital in Malaysia. Egypt J Otolaryngol. 2022;38: 135. doi: 10.1186/s43163-022-00331-w.

83. Nennstiel-Ratzel U, Arenz S, v. Kries R, Wildner M, Strutz J. [Modellprojekt Neugeborenen-Hörscreening in der Oberpfalz]. HNO. 2007;55: 128-134. doi: 10.1007/s00106-006-1383-x.

84. Olusanya BO, Wirz SL, Luxon LM. Hospital-based universal newborn hearing screening for early detection of permanent congenital hearing loss in Lagos, Nigeria. Int J Pediatr Otorhinolaryngol. 2008;72: 991-1001. doi: 10.1016/j.ijporl.2008.03.004.

85. Olusanya BO, Bamigboye BA. Is discordance in TEOAE and AABR outcomes predictable in newborns? Int J Pediatr Otorhinolaryngol. 2010;74: 1303-1309. doi: 10.1016/j.ijporl.2010.08.010.

86. Pasha YZ, Zamani M, Fard AH, Pasha EZ. Screening of hearing in newborn infants: follow-up and outcome after 40930 births in Babol, Northern Iran. Archives of Iranian medicine. 2018;21: 382-386.

87. Alothman N, Elbeltagy R, Mulla R. Universal newborn hearing screening program in Saudi Arabia: Current insight. J Otol. 2024;19: 35-39. doi: 10.1016/j.joto.2024.01.002.

88. Al Shamisi AM, Roy ME. Newborn Hearing Screening at a Single Tertiary Care Hospital in the United Arab Emirates. Aud Vestib Res. 2023;32: 54-63. doi: 10.18502/avr.v32i1.11322.

89. Ayas M, Yaseen H. Emerging Data from a Newborn Hearing Screening Program in Sharjah, United Arab Emirates. Int J Pediatr. 2021;2021: 2616890. doi: 10.1155/2021/2616890.

90. Clemens CJ, Davis SA. Minimizing false-positives in universal newborn hearing screening: a simple solution. Pediatrics. 2001;107: E29. doi: 10.1542/peds.107.3.e29.

91. Fan JY, Chen LS, Lai JC, Chen MK, Chen HC. A pre-paid newborn hearing screening programme: a community-based study. B-ENT. 2010;6: 265-269.

92. Gupta S, Sah S, Som T, Saksena M, Yadav CP, Sankar MJ, et al. Challenges of Implementing Universal Newborn Hearing Screening at a Tertiary Care Centre from India. Indian J Pediatr. 2015;82: 688-693. doi: 10.1007/s12098-015-1688-4.

93. Huang H, Chiang S, Shiau Y, Yeh W, Ho HC, Wang L, et al. The universal newborn hearing screening program of Taipei City. Int J Pediatr Otorhinolaryngol. 2013;77: 1734-1737. doi: 10.1016/j.ijporl.2013.08.004.

94. Iwasaki S, Hayashi Y, Seki A, Nagura M, Hashimoto Y, Oshima G, et al. A model of two-stage newborn hearing screening with automated auditory brainstem response. Int J Pediatr Otorhinolaryngol. 2003;67: 1099-1104. doi: 10.1016/s0165-5876(03)00199-x.

95. Kelly AF, Kelly PK, Shah M. Auditory Brainstem Response Pass Rates Correlate with Newborn Hour of Life and Delivery Mode. J Pediatr. 2021;230: 100-105. doi: 10.1016/j.jpeds.2020.10.036.

96. Messner AH, Price M, Kwast K, Gallagher K, Forte J. Volunteer-based universal newborn hearing screening program. Int J Pediatr Otorhinolaryngol. 2001;60: 123-130. doi: 10.1016/s0165-5876(01)00507-9.

97. Oruc MA, Alan Y, Mercan GC, Taner CE, Oncel MY, Alan M, et al. Results of newborn hearing screening in tepecik education and research hospital. Anatol J Family Med. 2021;4: 68-73. doi: 10.5505/anatoljfm.2021.83997.

98. Shim J, Kim H, Kwon Y, Chang J, Park E, Im GJ. Results of a 10-year hearing screening using automated auditory brainstem response in newborns: The two-step AABR method. Int J Pediatr Otorhinolaryngol. 2021;151: 110947. doi: 10.1016/j.ijporl.2021.110947.

99. Tanyeri Toker G, Kumbul YC, Cetinkol AE, Aslan H, Baba P, Oncel MY. Is Gestational COVID-19 a Risk Factor for Congenital Hearing Loss? Otol Neurotol. 2023;44: 115-120. doi: 10.1097/MAO.0000000000003761.

100. Tsuchiya H, Goto K, Yunohara N, Matsuoka M, Nishioka M, Nakamura Y, et al. Newborn hearing screening in a single private Japanese obstetric hospital. Pediatr Int. 2006;48: 604-607. doi: 10.1111/j.1442-200X.2006.02274.x.

101. Hewlett J, Waisbren SE. A review of the psychosocial effects of false-positive results on parents and current communication practices in newborn screening. J Inherit Metab Dis. 2006;29: 677-682. doi: 10.1007/s10545-006-0381-1.

102. Gemeinsamer Bundesausschuss. [Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Kinder-Richtlinien: Einführung eines Neugeborenen-Hörscreenings]. 2008. [cited Apr 3, 2024]. Available from: https://www.g-ba.de/downloads/40-268-641/2008-06-19-Kinder-H%C3%B6rscreening_TrG.pdf.

103. Hunter LL, Meinzen-Derr J, Wiley S, Horvath CL, Kothari R, Wexelblatt S. Influence of the WIC Program on Loss to Follow-up for Newborn Hearing Screening. Pediatrics. 2016;138: e20154301. doi: 10.1542/peds.2015-4301.