Changing the Paradigm for School Hearing Screening Globally: Evaluation of Screening Protocols From Two Randomized Trials in Rural Alaska

Samantha Kleindienst Robler,[1,2] Alyssa Platt,[3,4] Cole D. Jenson,[2] S. Meade Inglis,[1,3,5] Philip Hofstetter,[6] Alexandra A. Ross,[5,7] Nae-Yuh Wang,[8,9] Alain Labrique,[10] Joseph J. Gallo,[11] Joseph R. Egger,[3] and Susan D. Emmett[1,3,5,7,12]

Objectives: Diagnostic accuracy was evaluated for various screening tools, including mobile health (mHealth) pure-tone screening, tympanometry, distortion product otoacoustic emissions (DPOAE), and inclusion of high frequencies to determine the most accurate screening protocol for identifying children with hearing loss in rural Alaska where the prevalence of middle ear disease is high.

Design: Hearing screening data were collected as part of two cluster randomized trials conducted in 15 communities in rural northwest Alaska. All children enrolled in school from preschool to 12th grade were eligible. Analysis was limited to data collected 2018 to 2019 (n = 1449), when both trials were running and measurement of high frequencies were included in the protocols. Analyses included estimates of diagnostic accuracy for each screening tool, as well as exploring performance by age and grade. Multiple imputation was used to assess diagnostic accuracy in younger children, where missing data were more prevalent due to requirements for conditioned responses. The audiometric reference standard included otoscopy, tympanometry, and high frequencies to ensure detection of infection-related and noise-induced hearing loss.

Results: Both the mHealth pure-tone screen and DPOAE screen performed better when tympanometry was added to the protocol (increase in sensitivity of 19.9%, 95% Confidence Interval (CI): 15.9 to 24.1 for mHealth screen, 17.9%, 95% CI: 14.0 to 21.8 for high-frequency mHealth screen, and 10.4%, 95% CI: 7.5 to 13.9 for DPOAE). The addition of 6 kHz to the mHealth pure-tone screen provided an 8.7 percentage point improvement in sensitivity (95% CI: 6.5 to 11.3). Completeness of data for both the reference standard and the mHealth screening tool differed substantially by age, due to difficulty with behavioral testing in young children. By age 7, children were able to complete behavioral testing, and data indicated that high-frequency mHealth pure-tone screen with tympanometry was the superior tool for children 7 years and older. For children 3 to 6 years of age, DPOAE plus tympanometry performed the best, both for complete data and multiply imputed data, which better approximates accuracy for children with missing data.

Conclusions: This study directly evaluated pure-tone, DPOAE, and tympanometry tools as part of school hearing screening in rural Alaskan children (3 to 18+ years). Results from this study indicate that tympanometry is a key component in the hearing screening protocol, particularly in environments with higher prevalence of infection-related hearing loss. DPOAE is the preferred hearing screening tool when evaluating children younger than 7 years of age (below 2nd grade in the United States) due to the frequency of missing data with behavioral testing in this age group. For children 7 years and older, the addition of high frequencies to pure-tone screening increased the accuracy of screening, likely due to improved identification of hearing loss from noise exposure. The lack of a consistent reference standard in the literature makes comparing across studies challenging. In our study with a reference standard inclusive of otoscopy, tympanometry, and high frequencies, less than ideal sensitivities were found even for the most sensitive screening protocols, suggesting more investigation is necessary to ensure screening programs are appropriately identifying noise- and infection-related hearing loss in rural, low-resource settings.

Key words: Child health, Hearing loss, Hearing screening, Infection-related hearing loss, Otitis media, Otoacoustic emission screening, Pure-tone screening, Rural health, School screening, Tympanometry.

(Ear & Hearing 2023;XX;00–00)

IDEA

Inclusion, Diversity, Equity, Accessibility Article.

[1]Department of Otolaryngology, Head & Neck Surgery, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA; [2]Department of Audiology, Norton Sound Health Corporation, Nome, Alaska, USA; [3]Duke Global Health Institute, Durham, North Carolina, USA; [4]Department of Biostatistics & Bioinformatics, Duke School of Medicine, Durham, North Carolina, USA; [5]Center for Health Policy and Inequalities Research, Duke University, Durham, North Carolina, USA; [6]Petersburg Medical Center, Petersburg, Alaska, USA; [7]Department of Head and Neck Surgery and Communication Sciences, Duke University School of Medicine, Durham, North Carolina, USA; [8]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [9]Departments of Biostatistics and Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [10]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [11]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; and [12]Department of Epidemiology, Fay W. Boozman College of Public Health, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

INTRODUCTION

Seventy million children in the world have hearing loss, and underserved populations are disproportionately affected (Haile et al. 2021). The World Health Organization (WHO) estimates that 60% of childhood hearing loss is preventable (World Health Organization 2021), with this estimate rising to 75% in low-resource settings where infection-related hearing loss is common (Smith & Boss 2010; World Health Organization 2012; Olusanya et al. 2014). Childhood hearing loss has well-known, lifelong effects on educational attainment, psychosocial

<zdoi; 10.1097/AUD.0000000000001336>

outcomes, and vocational opportunities, and treatment significantly improves outcomes (Järvelin et al. 1997; Bess et al. 1998; Wake et al. 2004; Kennedy et al. 2006; Khairi Md Daud et al. 2010; Jung & Bhattacharyya 2012; Lieu et al. 2012; Emmett & Francis 2015; Tomblin et al. 2015). Consequently, early identification of childhood hearing loss is crucial for prevention and treatment, especially in low-resource settings (Robinshaw 1995; Moeller 2000).

School hearing screening is an essential public health approach to addressing childhood hearing loss (Flanary et al. 1999; Anderson et al. 2011; Swanepoel et al. 2013; World Health Organization 2020). Hearing screening is particularly important in rural, underserved regions, where access to care is limited and communities often experience a high burden of infection-related hearing loss. Rural Alaska is one such example, where nearly 75% of rural communities are not connected to a hospital by road and most physicians are concentrated in urban areas (Hofstetter et al. 2010; Carroll et al. 2011; Kokesh et al. 2011). While Alaska mandates school hearing screenings, recommendations for specific screening protocols are not included in the mandate, and the effectiveness of existing protocols has yet to be evaluated (Alaska Statutes 2019).

Despite consensus on the need for school hearing screening, screening guidelines are inconsistently implemented and often lack the necessary scientific rigor (Anderson et al. 2011; Skarzynski & Piotrowska 2012; Sekhar et al. 2013; Prieve et al. 2015; Yong et al. 2020). A recent review of school hearing screening programs globally by Yong et al. (2020) found the presence of screening programs to be inconsistent regardless of mandates, and screening protocols to be variable. The most common protocols included pure-tone screening, but specifics regarding which and number of frequencies were inconsistent, and the use of additional testing such as otoacoustic emissions (OAE), otoscopy and tympanometry varied. Pure-tone screening protocols most commonly included 0.5, 1, 2, 4 kHz, with recommendations to add 6 and 8 kHz in adolescents due to noise exposure (Sekhar et al. 2016). OAE screening was recommended only when indicated, such as with children who are unable to follow directions (<3 years of age; Anderson et al. 2011). Threshold definitions for follow-up referral also varied, with studies and guidelines ranging from 25 to 40 dB HL. Most concerningly, many studies do not include a benchmark audiometric assessment, which is needed to assess diagnostic accuracy and the true prevalence of hearing loss among school children (Yong et al. 2020). The establishment of standardized evidence-based school hearing screening guidelines is essential to improve existing screening programs, inform policy development, drive high quality research, and better measure the impact of screening and treatment interventions.

To begin to address this gap, the diagnostic accuracy of various hearing screening protocols was evaluated. Protocols included mobile health (mHealth) pure-tone screening, tympanometry, distortion product otoacoustic emissions (DPOAE), and inclusion of high frequencies against a benchmark audiometric evaluation for children, preschool to 12th grade. This evaluation was done as part of two cluster randomized trials in rural Northwest, Alaska that were designed to evaluate a new school hearing screening and follow-up process using mHealth and telemedicine solutions (Emmett et al. 2022). Our aim was to determine the most accurate screening protocol for identifying children with hearing loss in rural Alaska where the prevalence of middle ear disease is high.

MATERIALS AND METHODS

Study Design

Hearing Norton Sound comprised two cluster randomized controlled trials testing digital innovations to improve timely identification and treatment of childhood hearing loss in rural Alaska, with full protocols published previously (Emmett et al. 2019a, b). Briefly, the main trial was conducted over two academic years (2017 to 2019) in the 15 communities in rural Northwest Alaska served by the Bering Strait School District. In the second year of the trial, an ancillary trial was added to include preschool children in the 14 communities within the region that had preschools. All students enrolled in preschool and grades K–12 were invited to participate. Written consent and verbal child assent were obtained for all children, with parental consent obtained for participants younger than 18 years of age. All participating children underwent the school hearing screening protocol (otoacoustic emission screening), a mHealth plus tympanometry screening protocol, and a benchmark audiometric assessment (Fig. 1). A referral was generated if any protocol indicated the need for follow-up testing (in other words, did not pass). The benchmark audiometric assessment completed on screening day was not used as a formal diagnostic assessment but as a reference standard for evaluating the accuracy of the screening protocols. If abnormal results were found on the benchmark assessment, a referral was generated similar to the screening protocols, and the child entered one of the two referral pathways, which were cluster randomized at the community level.

The hearing screening protocols and audiometric assessment were conducted on the same day in each of the local schools and preschool centers. Testing occurred in quiet rooms available during school hours and included empty classrooms, libraries, or school offices. Noise levels were monitored using functionality available within the equipment. Rooms were marked for hearing testing, and quiet signs were hung to reduce external noise from students. Testing was paused during the school bell and noisier student movement activities. If the room became too noisy or equipment indicated a noisy environment, testing was paused until the noise stopped. School staff completed the school hearing screening, layperson study staff completed the mHealth plus tympanometry screening, and audiologists completed the benchmark audiometric assessment. All study team members and school staff were blinded to the results of the other testing protocols.

The Institutional Review Boards of Alaska Area, Norton Sound Health Corporation, and Duke University approved the trial. The trials are registered with ClinicalTrials.gov (NCT03309553, NCT03662256).

Equipment and Procedures

School Screening Protocol • In Bering Strait School District, students are screened annually using DPOAE (Biologic AudX, Natus, Denmark). All school hearing equipment was calibrated annually per American National Standards on Acoustics (ANSI) requirement S3.6. The screening was automated and involved the screener placing a soft tip in the ear and recording a pass or refer for each ear. If there was too much noise or if the tip

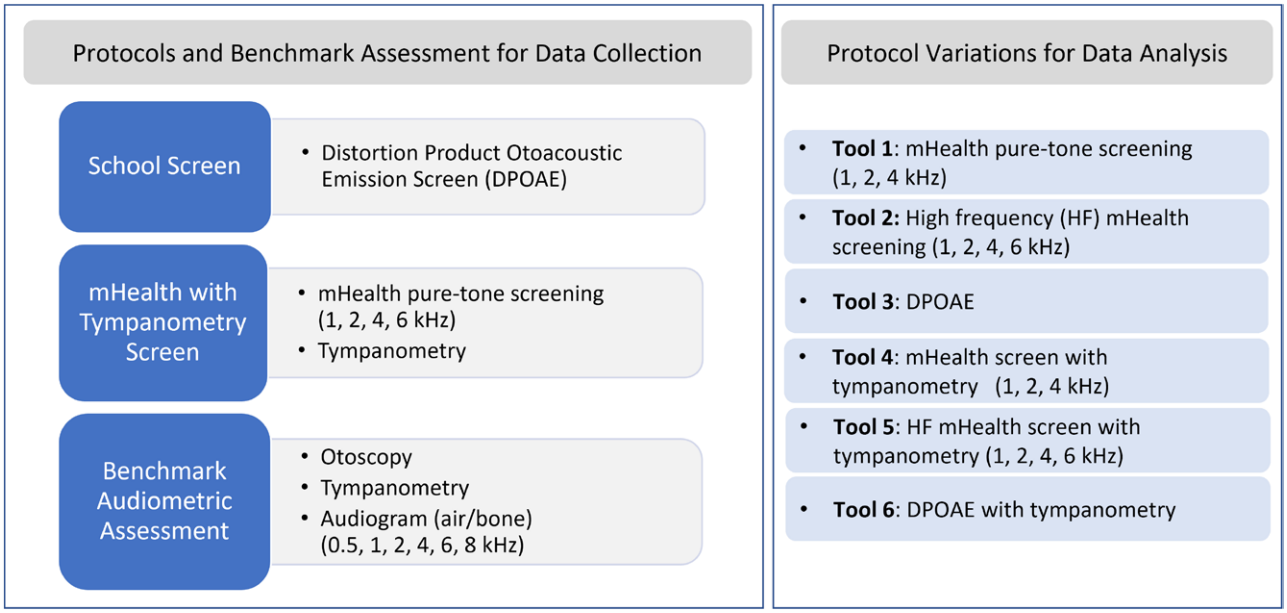

Fig. 1. Overview of protocols used for trial data collection and protocols evaluated during data analysis.

TABLE 1. Reference standard and index tool referral criteria*

Definitions of Target Conditions*	
Criteria	Definition
Middle ear disease	• Referral on otoscopy (retraction, effusion, acute otitis media, otorrhea, perforation, presence of tympanostomy tube, or external otitis) OR • Referral on tympanometry (type B or negative pressure <–200 daPa)
Audiometric evaluation with high frequency (>25 dB standard)	• Pure-tone average >25 dB (World Health Organization 2014) (0.5, 1, 2, 4 kHz) or an individual frequency ≥30 dB (0.5, 1, 2, 4, 6, 8 kHz); OR • Type B or C tympanogram (<–200 daPa) (FitzZaland & Zink 1984) OR Audiologist-interpreted pathology on otoscopy (retraction, effusion, acute otitis media, otorrhea, perforation, presence of tympanostomy tube, external otitis, cerumen impaction, foreign body)
Audiometric evaluation with high frequency (≥20 dB standard)	• Pure-tone average ≥20 dB (World Health Organization 2021) (0.5, 1, 2, 4 kHz) in either ear or an individual frequency ≥30 dB (0.5, 1, 2, 4, 6, 8 kHz); OR • Type B or C tympanogram (<–200 daPa) (FitzZaland & Zink 1984) OR • Audiologist-interpreted pathology on otoscopy (retraction, effusion, acute otitis media, otorrhea, perforation, presence of tympanostomy tube, external otitis, cerumen impaction, foreign body)
Index (Screening) Tool Referral Criteria*	
Evaluation	Criteria for Referral
mHealth pure-tone screen	• No response at 20 dB at 1, 2, or 4 kHz on rescreen (Mahomed-Asmail et al. 2016)
High-frequency mHealth pure-tone screen	• No response at 20 dB at 1, 2, 4, or 6 kHz on rescreen (Mahomed-Asmail et al. 2016)
School screen (DPOAE)	• 2 or more frequencies (2, 3, 4, 5 kHz) did not meet predetermined criteria for otoacoustic emission (Gorga et al. 1997)
mHealth pure-tone screen plus tympanometry	• No response at 20 dB at 1, 2, or 4 kHz on rescreen (Mahomed-Asmail et al. 2016) OR • Type B or C tympanogram (<–200 daPa) (FitzZaland & Zink 1984)
High-frequency mHealth pure-tone screen plus tympanometry	• No response at 20 dB at 1, 2, 4, or 6 kHz on rescreen (Mahomed-Asmail et al. 2016) OR • Type B or C tympanogram (<–200 daPa) (FitzZaland & Zink 1984)
DPOAE plus tympanometry	• 2 or more frequencies (2, 3, 4, 5 kHz) did not meet predetermined criteria for otoacoustic emission (Gorga et al. 1997) • Type B or C tympanogram (<–200 daPa) (FitzZaland & Zink 1984)

*All target condition definitions and referral criteria were prespecified.
DPOAE, distortion product otoacoustic emissions.

was occluded, the machine prompted the screener to address the issue. Emissions were measured at 2, 3, 4, and 5 kHz for each ear using an overall pass/refer criteria in which three out of four frequencies had to meet predetermined criteria (Gorga et al.

1997), with no formal rescreen process (Table 1). School teachers, typically special education teachers and support staff, completed the school screening. Training was provided to teachers by school administration (e.g., Director of Special Education Services), and technical support was provided by audiology staff at the Norton Sound Health Corporation, as is standard practice.

mHealth Plus Tympanometry Screening Protocol • The mHealth screen plus tympanometry protocol included a smartphone-based pure-tone hearing screening and a middle ear evaluation using tympanometry (FitzZaland & Zink 1984; Swanepoel et al. 2014; Mahomed-Asmail et al. 2016; Yousuf Hussein et al. 2016, 2018). Layperson study staff completing the screening received initial training by audiology study staff that included how to use the equipment, categorize tympanogram types, and basic troubleshooting. A protocol that screened both hearing and middle ear function was selected based on systematic reviews and practice guidelines (FitzZaland & Zink 1984; Prieve et al. 2015) for school screening, as well as consideration of the prevalence of infection-related hearing loss in rural Alaska (Reed et al. 1967; Curns et al. 2002; Singleton et al. 2009, 2018).

Pure-tone hearing screening was administered using a validated, smartphone-based hearing screening application (hearX HearScreen, South Africa). Before screening, each participant received verbal instruction and a practice session built into the software was performed to confirm their understanding of the task. Frequencies screened included 1, 2, 4, and 6 kHz, presented at 20 dB HL (with the additional higher frequency added in the second year of the trial; Sekhar et al. 2016). Ambient noise levels were monitored, and the screener was notified to pause testing if excessive levels were reached. A lay-friendly green, yellow, or red bar also continuously displayed on the screen throughout testing to alert the screener of noise levels. Any frequency that did not elicit a response was rescreened at the end of initial screening (Mahomed-Asmail et al. 2016). A referral was generated if there was no response to any frequency in either ear on the rescreen (Table 1).

Tympanometry was performed with a diagnostic tympanometer set to screening mode (Otoflex 100, Otometrics, Denmark). A referral was generated in the case of a flat (Type B) tympanogram or negative pressure <−200 daPa (Table 1; FitzZaland & Zink 1984; Lyons et al. 2004). All study equipment was calibrated annually per ANSI requirement S3.6.

Benchmark Audiometric Assessment • On hearing screening day, an audiometric assessment was performed on all participants to provide a reference standard for the screening protocols. This assessment included diagnostic audiometry using Shoebox, a validated mHealth tablet audiometer (Shoebox Audiometry, Clearwater Clinical, Canada; Thompson et al. 2015), tympanometry (Otometrics Otoflex 100, Denmark), and digital otoscopy (Otometrics Otocam, Denmark). For pure-tone testing (air and bone), participants were given verbal instructions to raise their hand when they heard a beep. A practice tone was given at 1000 Hz to confirm understanding of the task. Following standard practice in audiometric assessment, each threshold was obtained and then confirmed during testing to ensure reliability. Study team members who are trained audiologists performed the assessment. They were blinded to the results of the other screening protocols. Air conduction thresholds were assessed at 0.5, 1, 2, 4, 6, and 8 kHz (with the higher frequencies added in the second year of the trial). Referrals from diagnostic audiometry were generated for any single air conduction threshold ≥30 dB HL, or a pure-tone average (PTA) of >25 dB HL (0.5, 1, 2, and 4 kHz) (World Health Organization 2014). If maximum ambient noise levels were reached, a notification was sent to the audiologist who paused testing. Individual thresholds ≥30 dB HL were included in the referral criteria to capture middle ear disease and/or noise-induced hearing loss that may only affect one frequency. Bone conduction testing was performed if the corresponding air conduction threshold exceeded 25 dB HL.

Tympanometry and digital otoscopy were performed by the audiologist. Type B (flat) tympanograms or Type C (negative pressure <−200 daPa) generated a referral. Digital otoscopy was performed, with pathological findings (e.g., occluding cerumen, retraction, effusion, acute otitis media, otorrhea, perforation, patent or plugged tube, external otitis, foreign body) generating a referral at the discretion of the audiologist. See Table 1 for a summary of referral criteria for the screening protocols and benchmark audiometric assessment. All study equipment was calibrated annually per ANSI requirement S3.6.

Statistical Analysis

This study evaluated the diagnostic accuracy of 6 variations of the two screening protocols from the second year of the trial (Fig. 1 and Table 1):

1. Tool 1: mHealth pure-tone screening (1, 2, 4 kHz) alone
2. Tool 2: High-frequency mHealth pure-tone screening (1, 2, 4, 6 kHz)
3. Tool 3: DPOAE
4. Tool 4: mHealth pure-tone screening (1, 2, 4 kHz) with tympanometry
5. Tool 5: High-frequency mHealth pure-tone screening (1, 2, 4, 6 kHz) with tympanometry
6. Tool 6: DPOAE with tympanometry

This analysis included an estimation of the diagnostic accuracy of each of the six protocol variations, quantification of the differences between diagnostic accuracy of various protocols, and an exploration of differential accuracy by age group, defined by an evaluation of missing data (age 3 to 6 versus 7 years of age and older), and grade (preschool to 1st versus 2nd grade and up). Differences were explored by age due to the higher prevalence of middle ear disease and the difficulty of obtaining behavioral responses in the youngest children. Analyses were limited to data from the second year of the main trial due to the addition of the ancillary trial for preschool children and high-frequency data in year 2 (2018 to 2019).

In March of 2021, the WHO published new guidelines by which to judge the presence of hearing loss in children (World Health Organization 2021), specifying that the cutoff for hearing loss be lowered to a PTA of at least 20 dB HL from the previous version of a PTA greater than 25 dB HL (World Health Organization 2014). Both the new and former WHO definitions of hearing loss were included in this analysis.

Confusion matrices were presented for all index screening tools, with STARD diagrams constructed for each index protocol to illustrate sample flow from eligibility to formation of the analytic sample, consistent with STARD guidelines (Bossuyt et al. 2016).

Diagnostic accuracy statistics included calculations of sensitivity (Se), specificity (Sp), positive predictive value (PPV),

negative predictive value (NPV), Youden index, likelihood ratio for positive test results (LR+), likelihood ratio for negative test results (LR−), and percent concordance (overall accuracy). All diagnostic accuracy statistics were calculated at the level of the ear. Because accuracy statistics were computed at the level of the ear, clustering in outcomes (by child) is likely and was accounted for by calculating each statistic using generalized estimating equations (GEE; Leisenring et al. 1997; Rutter 2000; Genders et al. 2012). GEE regressions were specified with binomial distribution with identity link and an independence working correlation matrix. Details of the regression specification and calculation of each type of diagnostic accuracy statistic can be found in the Supplemental Digital Content, http://links.lww.com/EANDH/B85. Differences in mean estimates were calculated between tools, with 95% confidence intervals constructed using bootstrapping (Leisenring et al. 1997; Rutter 2000). See Supplemental Digital Content, http://links.lww.com/EANDH/B85, for further details on the methods.

Our benchmark assessment was defined using referral status based on a PTA greater than 25 dB HL on 0.5, 1, 2, and 4 kHz tones or findings of any individual tone of at least 30 dB HL on 0.5, 1, 2, 4, 6, or 8 kHz tones for audiometry (World Health Organization 2014), pathological findings on otoscopy, or Type B or C findings on tympanometry (see Table 1 for detailed definition). To be used as benchmark data for complete case analysis, at least three tones were required (0.5, 1, 2, or 4 kHz) to be nonmissing and at least two high-frequency tones (4, 6, or 8 kHz) to be nonmissing. In addition, both tympanometry and otoscopy were required to have nonmissing values.

Exploratory analyses were conducted using alternative reference standard definitions: (1) a benchmark audiometric evaluation with referral thresholds of ≥20 dB HL (current WHO definition for hearing loss); and (2) using middle ear disease as the referent condition. Diagnostic accuracy was also analyzed by age (3 to 6 years versus 7 years of age and older) by introducing interaction terms to our regressions to calculate stratified accuracy statistics.

The original study was not specifically powered to detect a particular precision for the calculation of diagnostic accuracy; thus, no justification for study sample size is presented.

Missing Data

Missing reference data were expected to be more prevalent in younger participants due to requirements for conditioned response process for audiometry (the reference standard). It is also plausible that children with a history of hearing loss and/or middle ear disease are more likely to experience cognitive delay, resulting in the inability to condition and provide reference data, which may lead to a bias in some diagnostic accuracy statistics (Se, Sp, PPV, NPV; Whiting et al. 2004; Naaktgeboren et al. 2016). Therefore, as a sensitivity analysis, a multiple imputation process was implemented using chained equations (White et al. 2011) to produce an alternative set of estimates for each diagnostic accuracy metric (Naaktgeboren et al. 2016). Two variations of the analysis with imputation were performed. First, accuracy statistics were calculated using a multiply imputed reference standard but assumed any missing index tool component had a "pass" status. This was to accurately reflect the ability of the screening tool to identify ear/hearing pathology in the field (e.g., inability to obtain behavioral responses for pure-tone testing). Second, multiple imputation was used for both reference standard and index tool to estimate what the accuracy of the screening tool might have been if there were no missing data. Details of the multiple imputation process can be found in the Supplemental Digital Content, http://links.lww.com/EANDH/B85.

All analyses were conducted using Stata SE version 17 software.

RESULTS

Baseline Characteristics of the Analytic Sample

In year 2 of the trial, 1449 children were screened (1034 rescreened from the first study year, 262 new K–12 participants in the second study year, and 153 preschool children; see Fig. 2), of whom 1318 had benchmark audiometric assessments (reference standard) in at least 1 ear. Participant flow for each of the index screening protocols can be found in Figures 1–6, Supplemental Digital Content, http://links.lww.com/EANDH/B86.

Baseline characteristics of children screened in year 2 can be found in Table 2. The median age for children in the sample was 10 years old, with 71.2% of the sample below age 13. Slightly more participants were male (n = 777, 54%), most were Alaska Native (n = 1389, 96%), and most had at least 1 caregiver with a high school diploma or GED (n = 1,347, 95%). Based on the audiometric assessment, a total of 147 (10%) children had a mild hearing loss or greater (PTA of >25 dB HL at 0.5, 1, 2, 4 kHz) in at least 1 ear, and 246 (17%) had middle ear disease. Overall, nearly 8% of children had both middle ear disease and hearing loss, while 9.2% had middle ear disease with no hearing loss, and 2.8% had hearing loss without middle ear disease. A comparison of baseline characteristics for children with and without missing reference standard data is provided in Table 1, Supplemental Digital Content, http://links.lww.com/EANDH/B89. Children with missing data were more likely to be younger and have middle ear disease. Though missing data were less prevalent for DPOAE and tympanometry screening components, ears that had missing data for one of the tools were more likely to have pathological findings on the other tool (Table 2, Supplemental Digital Content, http://links.lww.com/EANDH/B89).

Diagnostic Accuracy

Concordance between index tools and the audiometric assessment (reference standard) ranged from 83 to 87% for complete case data (Table 3). The mHealth screen plus tympanometry (MS + Tymp) and high-frequency mHealth screen with tympanometry (HF MS + Tymp) had the highest concordance (86.9 and 87.4%, respectively). See Table 3, Supplemental Digital Content, http://links.lww.com/EANDH/B89 for confusion matrices for each of the screening tools compared with the full audiometric evaluation with high frequencies.

All diagnostic accuracy measures were completed using complete case data and two variations of multiple imputation: (1) full imputation of both reference standard and index tools; and (2) imputation of reference standard with missing index tools set to "pass," to more accurately reflect the ability of the screening tool to identify ear/hearing pathology in the field (e.g., mimicking the real-world scenario where an inability to obtain behavioral responses for pure-tone testing results in

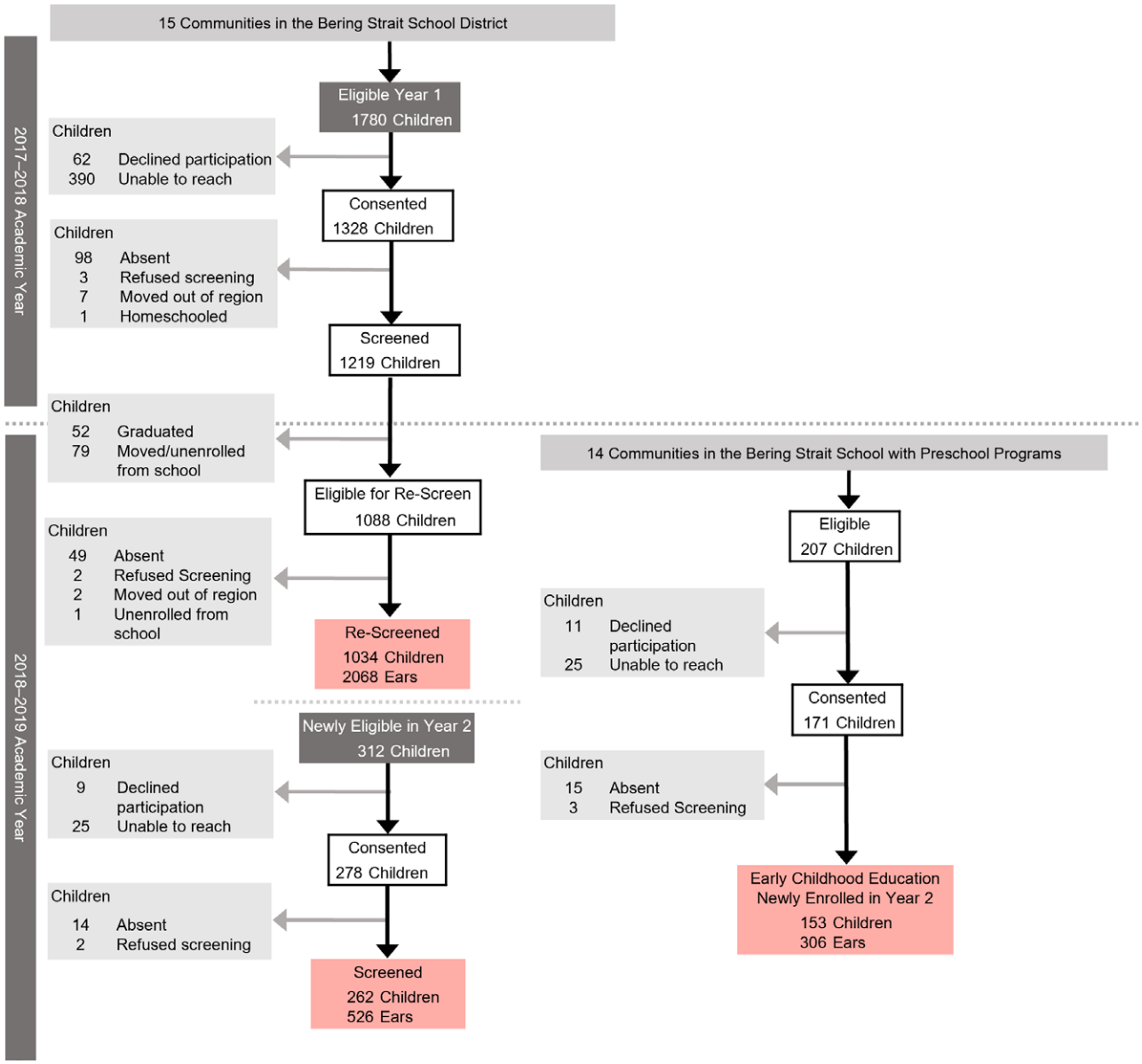

Fig. 2. STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) diagram for inclusion in study sample from main trial and ancillary trial. Final analytic study sample highlighted in red.

children not actually receiving a referral). Sensitivity, specificity, PPV, NPV, LR+, LR−, and Youden index are presented in Table 3 for complete case data and in Table 4, Supplemental Digital Content, http://links.lww.com/EANDH/B89, for the variations of multiple imputation. Results for each index tool are described later with a focus on the sensitivity of each tool in detecting hearing loss as defined by WHO (PTA of >25 dB HL) and/or middle ear disease as defined by pathological findings on otoscopy/tympanometry.

Diagnostic accuracy results were also evaluated by age using complete case and multiple imputation. It is well known that younger children may not be able to condition for behavioral audiometry. It is also known that younger children are more likely to have middle ear disease and infection-related hearing loss. For these reasons, it is expected to have missing data for reference and index tools that require behavioral responses. Figure 3 shows the proportion of children with missing data by age and grade for each component of the index tools (e.g., otoscopy, tympanometry, number of frequencies obtained). The proportion of children with any missing data for any ear was higher for children younger than 7 years of age (1st grade and below) compared with children ages 7 and older. As expected, the age-related pattern of missing data was most prominent in components of the reference standard and in index tools that required behavioral responses (e.g., pure-tone screening). Conversely, there was a low and consistent amount of missing data across all ages for objective measurements (e.g., DPOAE, tympanometry). Based on the observed cut point of increased missing data younger than age 7, diagnostic accuracy results were stratified by age group (3 to 6 years and 7 years and older) and are described for each of the index tools later. Diagnostic accuracy metrics stratified by age are displayed in Figure 4A, B and are described later for each index tool.

TABLE 2. Baseline sociodemographics and clinical characteristics of participants (N = 1449)

	All Screened, N (%)
Child age	
Mean (SD)	9.91 (4.08)
Median (Q1, Q3)	10.0 (7.0, 13.0)
Child age range	
3–6	346 (23.9%)
7–9	369 (25.5%)
10–12	317 (21.9%)
13–15	247 (17.1%)
16+	169 (11.7%)
Missing	1
Age at screening	
3–6	346 (23.9%)
7+	1102 (76.1%)
Missing	1
Female	671 (46.3%)
Missing	1
American Indian or Alaska Native	1389 (95.9%)
Grade level	
Preschool	153 (10.6%)
K–5	705 (48.7%)
6–8	304 (21.0%)
9–12	287 (19.8%)
Grade at screening	
Preschool–1st	383 (26.4%)
2nd–12th	1066 (73.6%)
Highest education level of any caregiver	
<12 grade	76 (5.3%)
HS diploma or GED	915 (64.3%)
Some college	290 (20.4%)
College degree	142 (10.0%)
Missing	26
HL severity	
No hearing loss in either ear	1244 (89.4%)
Mild (PTA >25–40 dB)	127 (9.1%)
Moderate+ (PTA 41+ dB)	20 (1.4%)
Missing	58
MED	
No	1193 (82.9%)
Yes	246 (17.1%)
Missing	10
Combined HL and MED status (>25 dB)	
No MED, no HL	1114 (80.3%)
MED, no HL	127 (9.2%)
HL, no MED	39 (2.8%)
MED + HL	107 (7.7%)
Missing	62

GED, general educational development; HL, Hearing loss; HS, high school; K, kindergarten; MED, Middle ear disease; PTA, pure-tone average.

Tool 1: mHealth Pure-Tone Screening (1, 2, 4 kHz) Alone • The sensitivity and specificity for the mhealth screening (MS) using complete case data was 40.3% (36.2 to 44.5) and 94.9% (93.9 to 95.9), respectively (Table 3). MS had the lowest sensitivity compared with the other index tools and the highest specificity, although specificity was greater than 90% for all the index tools. Overall, sensitivity and specificity of MS using multiple imputation for all children was similar to complete case findings (see Table 4, Supplemental Digital Content, http://links.lww.com/EANDH/B89). When evaluating diagnostic accuracy by age (Tables 4 and 5), sensitivity of MS dropped 17.9 percentage points in children 3 to 6 years of age compared with all children and was essentially unchanged for children 7 years of age and older. Using the fully imputed reference standard and index data to account for missing data in the youngest children (Fig. 4A), results suggested slightly higher sensitivity for children 3 to 6 years of age (orange open circles) than for older children. However, when the missing index tool was set to "pass," the sensitivity went down (open orange triangles), indicating lower sensitivity in real-world settings. When evaluating test performance by age, a heatmap was used to assess differences in sensitivity, specificity, and Youden index for the two age groups using complete case data (Fig. 5A, B; see Figures 7–10, Supplemental Digital Content, http://links.lww.com/EANDH/B87, for multiply imputed data), with MS showing the lowest performance. For middle ear disease, MS was the least sensitive (45.7%, 40.4 to 51.0) compared with the other index tools using complete case data (Table 3).

Tool 2: High-Frequency mHealth Pure-Tone Screening (1, 2, 4, 6 kHz) • The sensitivity and specificity for the high-frequency MS (HF MS) using complete case data was 49.1% (44.8 to 53.3) and 93.7% (92.6 to 94.8), respectively (Table 3) and were similar using multiple imputation (see Table 4, Supplemental Digital Content, http://links.lww.com/EANDH/B89). When evaluating diagnostic accuracy by age (Tables 4 and 5), sensitivity of HF MS dropped 26.7 percentage points in children 3 to 6 years of age compared with all children and remained unchanged for children 7 years of age and older. Similar to MS, when using the fully imputed reference standard and index data to account for missing data in the youngest children (Fig. 4A), results for HF MS suggested slightly higher sensitivity for children 3 to 6 years of age (orange open circles) than for older children, but when the missing index tool was set to "pass," the sensitivity dropped (open orange triangles).

Tool 3: DPOAE • The sensitivity and specificity for DPOAE screening using complete case data was 57.7% (53.5 to 61.8) and 91.3% (90.1 to 92.6), respectively (Table 3). When evaluating diagnostic accuracy by age (Tables 4 and 5), sensitivity was essentially the same for children 3 to 6 years of age and children 7 years of age and older compared with all children. For detecting middle ear disease, DPOAE was more sensitive (67.6%, 62.8 to 72.4) than MS or HF MS but not as sensitive as index tools that included tympanometry (Table 3).

Tool 4: mHealth Pure-Tone Screening (1, 2, 4 kHz) With Tympanometry • The sensitivity and specificity for the mhealth pure-tone screening with tympanometry (MS + Tymp) using complete case data was 60.2% (56.0 to 64.4) and 94.0% (92.9 to 95.0), respectively (Table 3) and was similar using multiple imputation (see Table 4, Supplemental Digital Content, http://links.lww.com/EANDH/B89). When evaluating diagnostic accuracy by age (Tables 4 and 5), sensitivity of MS + Tymp increased by 6.5 percentage points in children 3 to 6 years of age compared with all children and remained essentially unchanged for children 7 years of age and older. The primary driver of the increase in sensitivity of MS + Tymp for children 3 to 6 years of age was tympanometry, with little gained from using multiple imputation to address missing behavioral data in the youngest children (Fig. 4A, B).

Tool 5: High-Frequency mHealth Pure-Tone Screening (1, 2, 4, 6 kHz) With Tympanometry • The sensitivity and specificity for the high-frequency mhealth pure-tone screening with tympanometry (HF MS + Tymp) using complete case data was 67.0% (63.0 to 71.0) and 92.8% (91.6 to 93.9), respectively

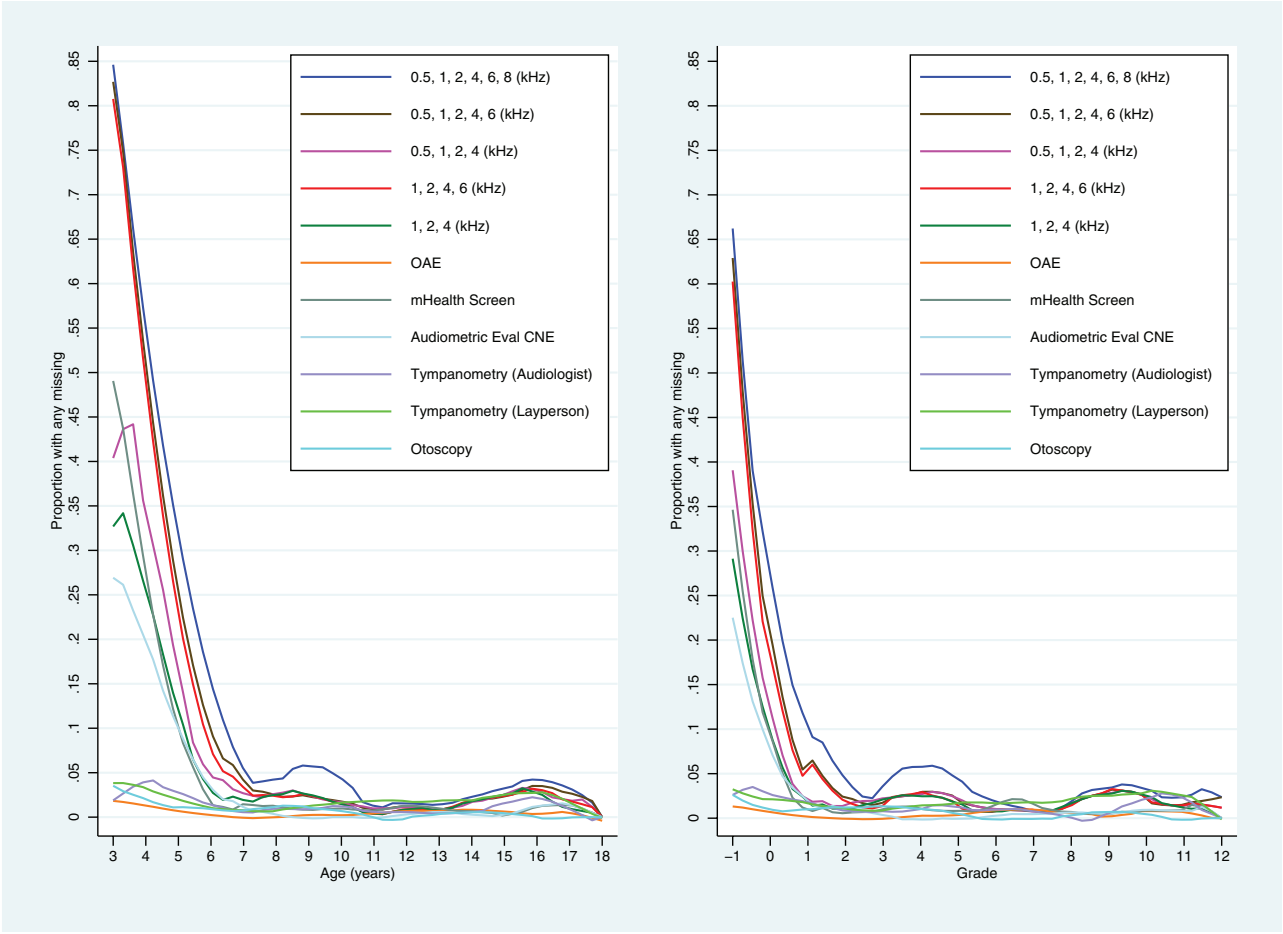

Fig. 3. Proportion of children with any data missing for various screening and evaluation components. CNE indicates could not evaluate; OAE, otoacoustic emissions.

(Table 3). The HF MS + Tymp had one of the two highest combined sensitivities and specificities (Youden index = 59.8%, 55.6 to 63.9), along with DPOAE plus tympanometry (DPOAE + Tymp; Youden index = 58.6%, 54.5 to 62.8), compared with the other index tools. Sensitivity and specificity of HF MS + Tymp using complete case data were similar using multiple imputation (see Table 4, Supplemental Digital Content, http://links.lww.com/EANDH/B89).

When evaluating diagnostic accuracy and test performance by age (Tables 4 and 5), sensitivity of HF MS + Tymp was essentially the same for children 3 to 6 years of age and children 7 years of age and older compared with all children. For children 7 years of age and older, complete case results (solid blue circles) suggest HF MS + Tymp had the highest sensitivity at 67.0% (62.7 to 71.3; Table 5), though differences between DPOAE + Tymp and high-frequency mHealth plus tympanometry were within the bounds of random variation (difference in sensitivity = 0.4%, −3.6 to 4.6; Fig. 5A). This was also seen on Youden index, which combines sensitivity and specificity together (Fig. 4B); performance was slightly better for HF MS + Tymp compared with DPOAE + Tymp in children 7 years of age and older (Youden index = 59.8%, 55.3 to 64.3; Fig. 4B; solid blue circles) compared with DPOAE + Tymp (difference = 3.4%, −1.1 to 7.6; Fig. 5B; solid blue circles).

Tool 6: DPOAE With Tympanometry • The sensitivity and specificity for DPOAE + Tymp screening using complete case data was 68.1% (64.1 to 72.1) and 90.5% (89.3 to 91.8), respectively (Table 3). Similar to HF MS + Tymp, DPOAE + Tymp yielded one of the highest sensitivities and specificities (Youden index = 58.6%, 54.5 to 62.8) compared with the other index tools and results were similar using multiple imputation (see Table 4, Supplemental Digital Content, http://links.lww.com/EANDH/B89).

When evaluating diagnostic accuracy by age using complete case data (Tables 4 and 5), sensitivity of the DPOAE + Tymp performed 10.2 percentage points better in children 3 to 6 years of age (Table 4) compared with all children (Table 3) and was essentially unchanged for children 7 years of age and older (Table 5). DPOAE + Tymp was the most sensitive index test for children 3 to 6 years old (Se = 78.3%, 68.5 to 88.0; Table 4), with 11.6 percentage point (3.7 to 21.0) higher sensitivity than the next most sensitive instruments (MS + Tymp and HF MS + Tymp; Fig. 5A, heatmap). For children 3 to 6 years old, DPOAE alone contributed 55.7% in sensitivity (41.1 to 67.4), with tympanometry providing the additional 22.5% (11.5 to 34.0) sensitivity (Fig. 5A). For comparison, tympanometry contributed 44.3% in sensitivity (28.8 to 57.4) when combined with HF MS in the youngest children. For older children (Fig. 5A; upper right panel), tympanometry performed similarly with the HF MS and DPOAE index tools, adding 14.2% (10.5 to 17.7) and 13.7% (8.8 to 19.8), respectively, suggesting that tympanometry is driving sensitivity for younger children due to worse performance with behavioral testing.

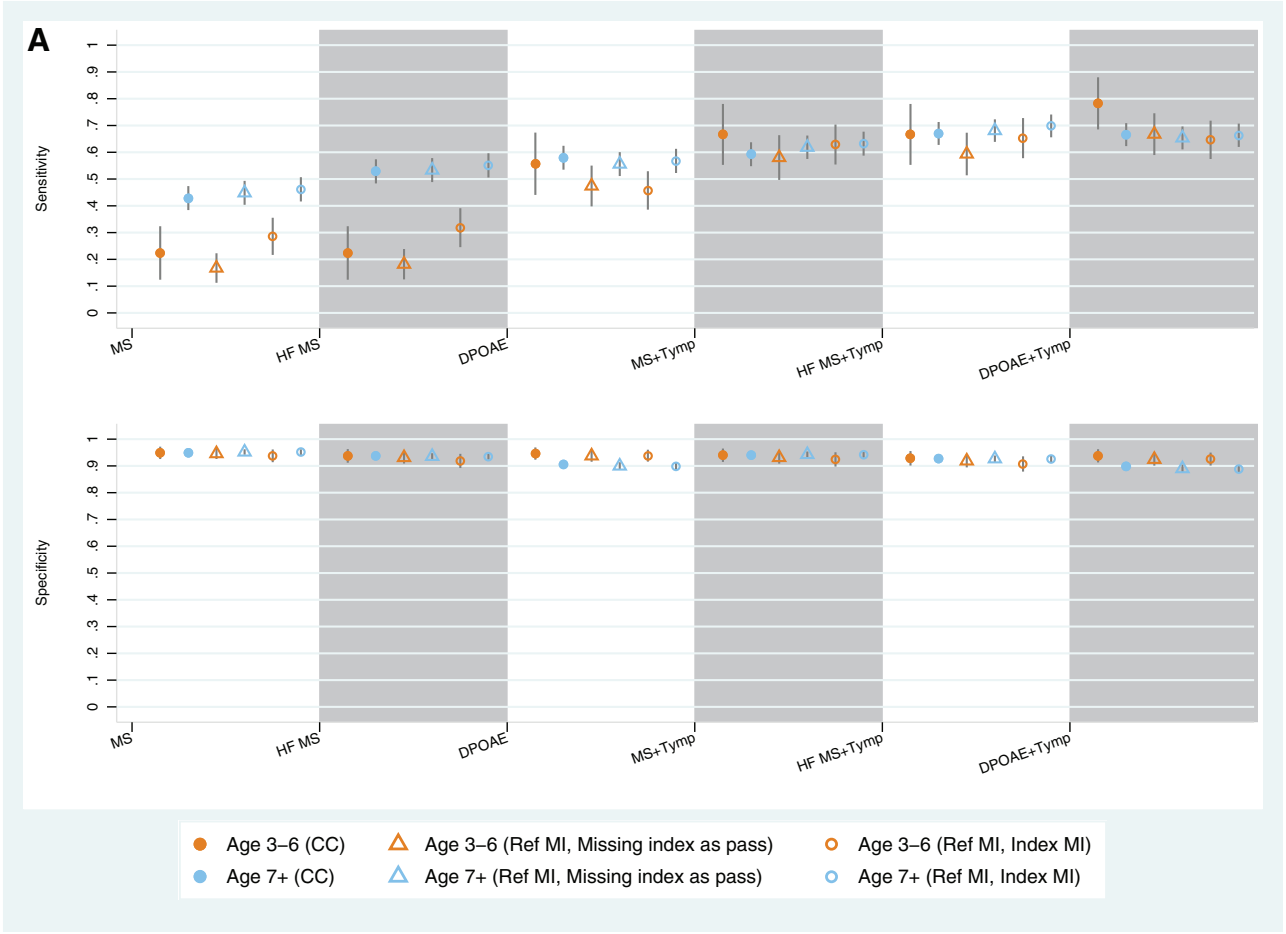

Fig. 4. Graphical representation of accuracy metrics of screening tools. A, Sensitivity, specificity of index tools with full audiometric evaluation (with high frequency) gold standard (PTA >25 dB criteria), stratified by age (3 to 6 y, 7+ y), based on complete case analysis (solid circles), multiply imputed reference standard (with missing index tools as "pass"), and fully multiply imputed data (missing reference and index tools imputed). B, Youden index, positive and negative likelihood ratios of index tools with full audiometric evaluation (with high frequency) gold standard (PTA >25 dB criteria), stratified by age (3 to 6 y, 7+ y), based on complete case analysis (full circles), multiply imputed reference standard (with missing index tools as "pass"), and fully multiply imputed data (missing reference and index tools imputed). CC indicates complete case; DPOAE, distortion product otoacoustic emissions; DPOAE + tymp, distortion product otoacoustic emissions with tympanometry; MS, mHealth screen; HF MS, high frequency mHealth screen; HF MS + Tymp, high frequency mHealth screen plus tympanometry; MI, multiple imputation; MS + Tymp, mHealth screen plus tympanometry; PTA, pure-tone average.

For children 3 to 6 years old, higher performance for the DPOAE + Tymp was also seen in Youden index, which combines sensitivity and specificity (Youden index = 72.0%, 62.0 to 82.1; Fig. 4B; solid orange circles), compared with HF MS + Tymp (difference = 12.5%, 3.7 to 22.8; Fig. 5B, heatmap). The reverse was true for older children, with HF MS + Tymp having the highest Youden index (59.9%, 55.3 to 64.3) compared with DPOAE + Tymp (Fig. 4B; solid blue circles; difference = 3.4%, −1.1 to 7.6).

For the detection of middle ear disease, DPOAE + Tymp had the highest sensitivity (88.2%, 84.8 to 91.5) compared with all the other index tests using complete case data (Table 3).

Diagnostic Accuracy by Grade

Diagnostic accuracy results compared by grade were similar to results stratified by age. See Tables 5–6, Supplemental Digital Content, http://links.lww.com/EANDH/B89, and Figures 11–14, Supplemental Digital Content, http://links.lww.com/EANDH/B88, for results presented by grade, which are grouped preschool to first grade (multiply imputed) and second grade and up (complete case).

Diagnostic Accuracy by Threshold Definition

The WHO established a new definition of hearing loss after the completion of this study, changing from a PTA of > 25 dB HL at 0.5, 1, 2, 4 kHz to ≥20 dB HL (a 10 dB difference; World Health Organization 2021). Despite using the former WHO definition of hearing loss in the study protocol, diagnostic accuracy by index tool was also evaluated using the new WHO definition for children 7 years of age and older. This analysis was limited to older children because more specific threshold data were obtained for older children (e.g., below 20 dB HL) compared with younger children, where often if 20 dB HL response was obtained (normal per study protocol), additional threshold testing was not done in these more difficult to test children. The pattern of results for the older children using the new ≥20 dB HL definition paralleled findings using the 25 dB HL definition, with the mHealth screen plus tympanometry and DPOAE plus tympanometry demonstrating the highest sensitivity (Table 5), and differences between the two falling within the range of random variation (difference = 0.2%, −3.9 to 4.4). However, overall sensitivity was reduced by about 5 to 10 percentage points across the index tools.

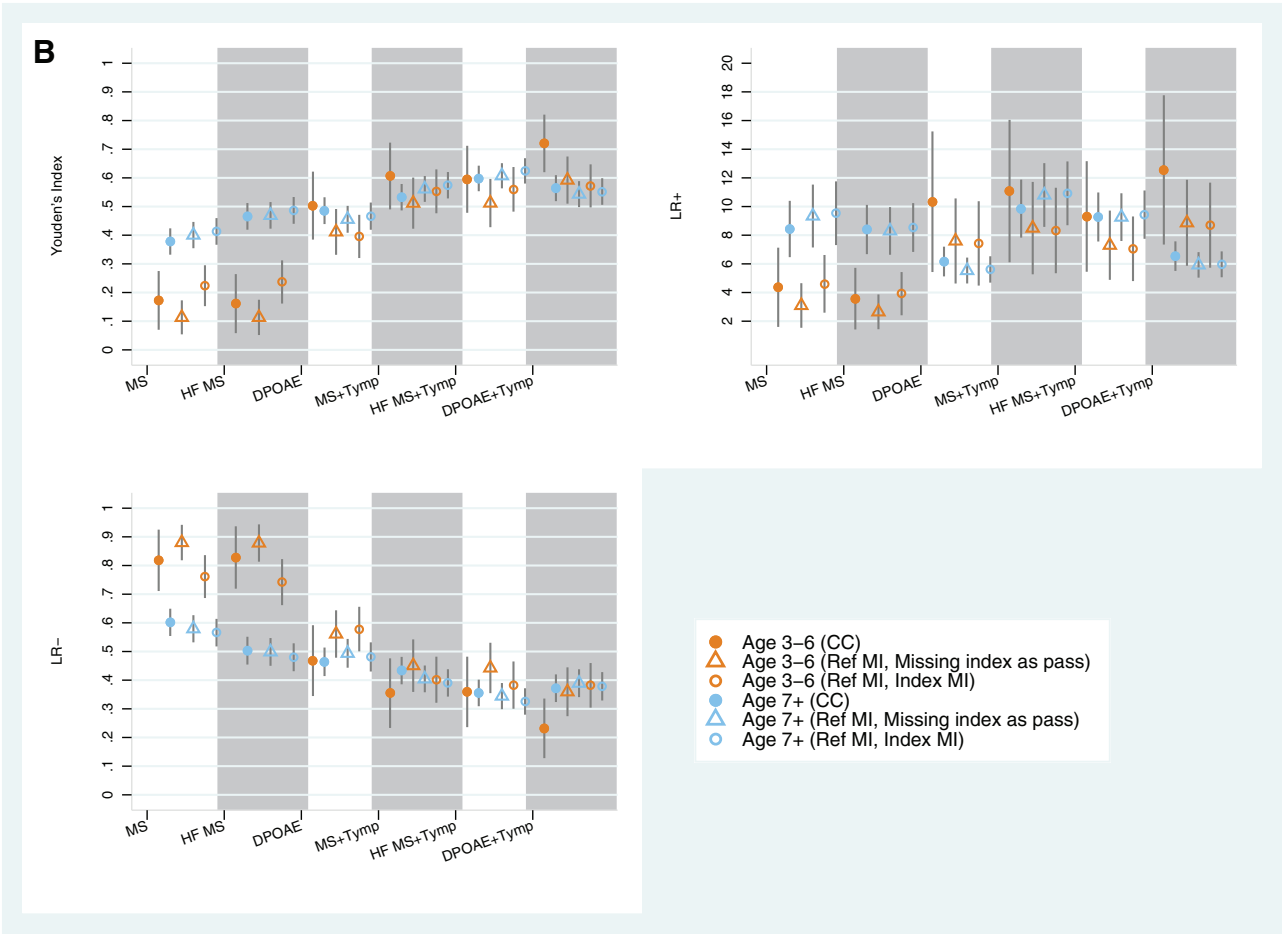

Fig. 4. Continued.

DISCUSSION

To our knowledge, this study includes the largest dataset that directly evaluated both behavioral pure-tone screening, objective OAE screening, and the inclusion of tympanometry for screening a large age range of children (3 to 18+ years of age) in a rural underserved area. Findings from this study, conducted in a rural Alaskan population, highlight the importance of incorporating tympanometry into screening protocols in rural environments with high prevalence of infection-related hearing loss. Findings also suggest the importance of adding high frequencies to the pure-tone screening to ensure identification of noise-induced hearing loss. Our data indicate a clear age cutoff (younger than 7 years) when OAEs are the preferred screening tool due to the degree of missing behavioral screening data among younger children who cannot reliably complete pure-tone testing.

The accuracy of two hearing screening protocols and their various components (mHealth pure-tone screening, DPOAE, tympanometry, and addition of high frequencies) were evaluated against a benchmark audiometric evaluation that included otoscopy, tympanometry, and pure-tone hearing testing. The mHealth pure-tone screen with high-frequency plus tympanometry (HF MS + Tymp) performed the best in children 7 years of age and older (2nd grade and up) and was more accurate than the school screening (DPOAE alone). In contrast, the DPOAE plus tympanometry (DPOAE + Tymp) protocol performed strongest in the youngest children (3 to 6 years of age), an age group that had difficulty completing the behavioral hearing screening and assessment. Despite limited evidence in the literature supporting the addition of tympanometry to the hearing screening protocol (Roberts 1976; FitzZaland & Zink 1984; Lyons et al. 2004), adding tympanometry to either DPOAE or mHealth pure-tone screening notably improved sensitivity in this rural Alaskan population where 9.2% of children had middle ear disease without hearing loss (PTA > 25 dB).

In general, sensitivity to hearing loss ranged from 40 to 68% for all variations of the screening protocols, which is lower than most of the previous literature (Prieve et al. 2015). However, the reference standard used, testing environment (sound booth versus field testing), and screening protocols in prior studies have all varied. A benchmark audiometric assessment that was inclusive of otoscopy and tympanometry was used. The comprehensive nature of our benchmark assessment may have contributed to the mHealth pure-tone screen alone yielding the lowest sensitivity (40.3%) in our analysis, which is lower than the 75% sensitivity previously reported using the same mHealth pure-tone screening in 1070 school-aged children (5 to 12 years of age; Mahomed-Asmail et al. 2016). There is an overall lack of consensus on what the reference standard should be, with most studies using various definitions of pure-tone testing (Prieve et al. 2015). Our comprehensive reference standard was used to ensure screening protocols are

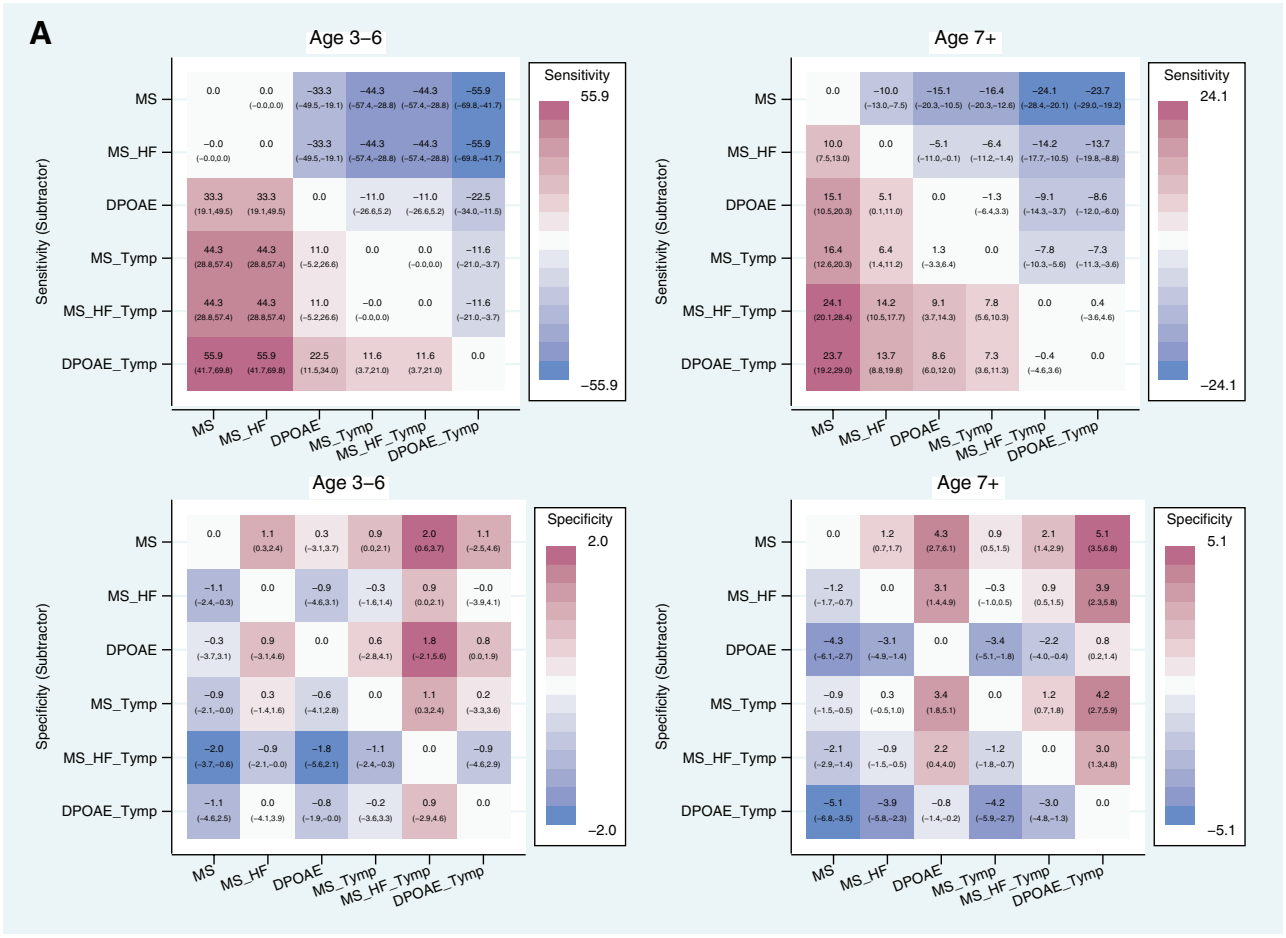

Fig. 5. Heatmap of differences between index tools. A, Heatmap of differences (with 95% confidence intervals) in sensitivity, specificity between index tools, using a full audiometric evaluation including high-frequency tones as gold standard (>25 dB criteria), with estimates stratified by age group (3 to 6, 7+) using complete case data and cluster bootstrapping with percentile confidence intervals. B, Heatmap of differences (with 95% confidence intervals) in Youden index between index tools, using a full audiometric evaluation including high-frequency tones as gold standard (>25 dB criteria), with estimates stratified by age group (3 to 6, 7+) using complete case data and cluster bootstrapping with percentile confidence intervals. DPOAE indicates distortion production otoacoustic emissions; DPOAE_Tymp, distortion product otoacoustic emissions with tympanometry; MS, mHealth screen; MS HF, High frequency mHealth screen; MS_tymp, mhealth Screen with tympanometry; MS_HF_Tymp, high frequency mHealth screen with tympanometry.

detecting children with hearing loss and/or middle ear disease in populations with high prevalence of infection-related hearing loss.

The addition of the 6 kHz high frequency to the mHealth pure-tone screen added 8.8 percentage points to sensitivity. Sekhar et al. (2016) found that adding 6 and 8 kHz to the standard pure-tone screening protocol of 0.5, 1, 2, and 4 kHz resulted in a 20-point jump in sensitivity from 58.1% (42.1 to 73.0) to 79.1% (64.0 to 90.0; p value = 0.003) and a nearly 19-point jump when adding 6 kHz only (76.7%, 61.4 to 88.2; p value 0.005). Conversely, the authors found a corresponding 10-point drop in specificity with the addition of high frequencies (91.2%, 83.4 to 96.1 to 81.3%, 71.8 to 88.7 for 6 and 8 kHz and 84.6%, 75.5 to 91.3 for 6 kHz), while only a 1.2 dB drop in specificity was found when adding 6 kHz (94.9%, 93.9 to 95.9 to 93.7%, 92.6 to 94.8). In our study, adding 6 kHz to the hearing screening increased sensitivity while preserving specificity.

The DPOAE screening alone had a sensitivity of 57.7% for hearing loss and performed better than mHealth pure-tone screening alone when evaluated across all children in our study. Few studies have directly compared OAE and pure-tone screening in the school environment (Sabo et al. 2000; Sliwa et al. 2011). These two studies looked at age ranges of 5 to 9 years (Sabo et al. 2000) and 10 to 14 years (Sliwa et al. 2011) and had variable results, which ultimately suggested that pure-tone screening performed better than OAE. However, pure-tone audiometry requires children to provide reliable behavioral responses. For younger children or for those who are difficult to test, pure-tone screening is often not possible. Children younger than 7 years of age (1st grade and below) were less likely to be able to complete behavioral testing, and thus, DPOAE was the more reliable and accurate assessment of hearing loss. This was evident in our data when evaluating the performance of each of the screening tools by two age groups: younger children (3 to 6 years) and older children (7 years of age and older).

To estimate performance in younger children where there were missing data due to the inability for younger children to compete behavioral responses, alternative estimates were obtained using two methods to account for missing data: one that estimated all missing reference standard data based on other available information (assuming one could obtain results for the index tool) and another that approximated real-world

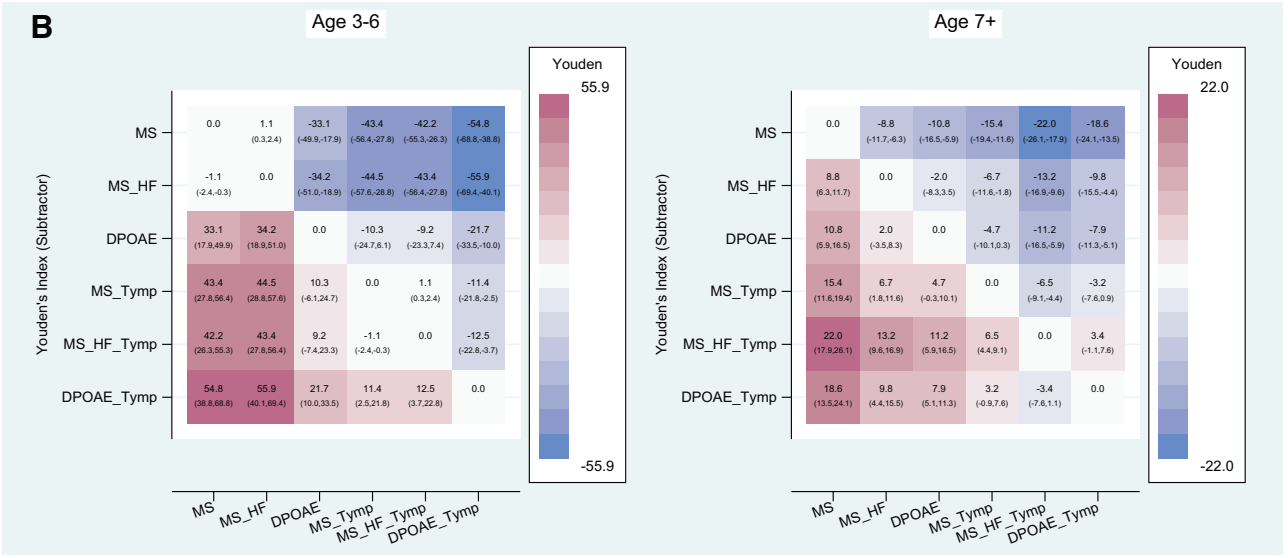

Fig. 5. Continued.

accuracy by setting the missing index results to "pass" (e.g., treating missing data as a failure in the ability to obtain a "refer" status in some children). While sensitivity estimates for the mHealth pure-tone screen increased for young children when using the full imputation analysis (from 22.4 to 28.6%), sensitivity decreased to 16.8% when setting the missing screening results to "pass," which is thought to closer approximate how the tool might perform in the real world in children from 3 to 6 years of age. Ultimately, these low sensitivities, even with statistical estimates for how these index tools would perform if results were obtained, suggest that behavioral hearing screening measures in children 3 to 6 years of age are not appropriate, and objective measures, such as OAE, are necessary.

The addition of tympanometry to the DPOAE and the mHealth pure-tone screen resulted in a 10- and 18-point improvement in sensitivity, respectively. Our results are consistent with those of Lyons et al. (2004), who found that the addition of tympanometry enhanced hearing screening performance in a sample of 1003 children 5 to 6 years old. The prevalence of middle ear disease in their sample was 13.2%, while our total sample (n = 1449, preschool to 12th grade) had a prevalence of middle ear disease of 17.1%. Among the children referred by tympanometry in Lyons' et al study, 65% passed pure-tone screening (pure-tone threshold >25 dB HL at 0.5, 1, 2, 4 kHz in at least 1 ear), corroborating our conclusion of the essential role of tympanometry in detecting middle ear disease in screening programs.

Finally, the importance of the definition of hearing loss for pure-tone screening and the implications for diagnostic performance should not be overlooked. The WHO recently published new guidelines that lower the definition for hearing loss to a PTA of ≥20 dB HL from the previous definition of a PTA >25 dB HL: a difference of 10 dB. Lowering the cutoff value, in theory, would reduce the number of children missed. However, school screening becomes logistically difficult at these levels due to the lack of sound-treated rooms (Driscoll et al. 2001; Lyons et al. 2004). Furthermore, a change in definition has an impact on screening protocols. At the time of our study, the WHO definition of hearing loss was PTA >25 dB. To maintain consistency with the study protocol but also report data that are comparable for future work, the performance of both the new and former WHO definitions in older children were evaluated. However, there are some inherent weaknesses in using the new definition with data from our study. In the youngest children, who were difficult to condition and fatigued quickly, study audiologists did not test below normal hearing as defined at the time. In other words, reliable responses at 20 dB HL were marked as threshold and considered normal hearing; however, using the new definition, these potentially suprathreshold responses could be artificially inflating the prevalence of hearing loss for the youngest children. Thus, a comparison using the new definition was feasible only for older children in our study and not possible for younger children.

The new WHO definition for hearing loss of ≥20 dB HL decreased the overall sensitivity across the index tools by 5 to 10 dB, with the tool with the highest sensitivity (HF MS + Tymp) decreasing by about 10 percentage points when compared with the 25 dB HL definition. This is due to both the protocol used for our study at the time, as well as the limits of the screening tools used. For the mHealth pure-tone screen, children were screened at 20 dB, with a positive response at all frequencies resulting in a pass. The screening protocol would need to be changed to screen at 15 dB HL across frequencies to accurately screen using the new WHO definition of hearing loss, and this low threshold may pose challenges with reliability in the absence of a sound-treated environment. The new WHO definition exceeds the limits of the DPOAE screening tool, where performance is best when normal hearing is defined as audiometric thresholds between 20 and 30 dB HL (Gorga et al. 1997). One can expect to see performance of both tools decrease with the more stringent definition for hearing loss. Future research in school-based hearing screenings will need to take the new WHO definition of hearing loss into consideration and accommodate for ambient noise experienced in the field (e.g., ANSI-standard headphones).

Limitations

There were significant missing data for younger children who could not perform behavioral pure-tone testing. Thus, multiple

TABLE 3. Diagnostic accuracy (with 95% confidence intervals) of index screening protocols by reference standard using complete case data for all ages

Tool	N	Sensitivity	Specificity	PPV	NPV	LR+	LR–	Youden Index	Concordance
Full audiometric evaluation with high frequency (>25 dB)*									
MS	2541	40.3 (36.2, 44.5)	94.9 (93.9, 95.9)	68.0 (62.9, 73.1)	85.6 (84.1, 87.0)	7.92 (6.22, 9.63)	0.63 (0.58, 0.67)	35.2 (31.0, 39.5)	83.4 (81.9, 84.8)
HF MS	2541	49.1 (44.8, 53.3)	93.7 (92.6, 94.8)	67.7 (63.1, 72.3)	87.3 (85.9, 88.7)	7.80 (6.32, 9.28)	0.54 (0.50, 0.59)	42.8 (38.4, 47.1)	84.3 (82.8, 85.7)
DPOAE	2560	57.7 (53.5, 61.8)	91.3 (90.1, 92.6)	64.1 (59.8, 68.3)	89.0 (87.6, 90.3)	6.65 (5.60, 7.71)	0.46 (0.42, 0.51)	49.0 (44.7, 53.3)	84.2 (82.8, 85.6)
MS + Tymp	2525	60.2 (56.0, 64.4)	94.0 (92.9, 95.0)	72.7 (68.5, 76.8)	89.9 (88.6, 91.2)	10.01 (8.14, 11.88)	0.42 (0.38, 0.47)	54.2 (49.9, 58.5)	86.9 (85.6, 88.2)
HF MS + Tymp	2525	67.0 (63.0, 71.0)	92.8 (91.6, 93.9)	71.1 (67.2, 75.1)	91.4 (90.1, 92.6)	9.28 (7.72, 10.84)	0.36 (0.31, 0.40)	59.8 (55.6, 63.9)	87.4 (86.1, 88.7)
DPOAE + Tymp	2542	68.1 (64.1, 72.1)	90.5 (89.3, 91.8)	65.6 (61.7, 69.6)	91.5 (90.2, 92.7)	7.20 (6.14, 8.26)	0.35 (0.31, 0.40)	58.6 (54.5, 62.8)	85.8 (84.5, 87.2)
Middle ear disease†									
MS	2733	45.7 (40.4, 51.0)	91.6 (90.5, 92.7)	43.5 (38.4, 48.7)	92.3 (91.2, 93.3)	5.45 (4.49, 6.40)	0.59 (0.53, 0.65)	37.3 (31.9, 42.7)	85.9 (84.6, 87.2)
HF MS	2733	50.1 (44.8, 55.5)	88.9 (87.7, 90.2)	39.1 (34.5, 43.7)	92.6 (91.6, 93.7)	4.53 (3.83, 5.23)	0.56 (0.50, 0.62)	39.1 (33.6, 44.5)	84.1 (82.7, 85.5)
DPOAE	2840	67.6 (62.8, 72.4)	87.5 (86.2, 88.8)	44.6 (40.5, 48.7)	94.8 (93.9, 95.7)	5.43 (4.74, 6.11)	0.37 (0.32, 0.43)	55.1 (50.2, 60.1)	85.0 (83.7, 86.3)
MS + Tymp	2711	82.6 (78.5, 86.7)	90.4 (89.3, 91.6)	54.2 (49.8, 58.6)	97.4 (96.8, 98.1)	8.63 (7.48, 9.78)	0.19 (0.15, 0.24)	73.0 (68.7, 77.3)	89.5 (88.3, 90.6)
HF MS + Tymp	2711	84.4 (80.5, 88.3)	87.9 (86.6, 89.2)	48.8 (44.7, 53.0)	97.6 (97.0, 98.3)	6.96 (6.14, 7.78)	0.18 (0.13, 0.22)	72.3 (68.1, 76.4)	87.5 (86.2, 88.7)
DPOAE + Tymp	2817	88.2 (84.8, 91.5)	86.6 (85.3, 87.9)	48.7 (44.8, 52.5)	98.1 (97.5, 98.6)	6.58 (5.87, 7.28)	0.14 (0.10, 0.18)	74.8 (71.1, 78.4)	86.8 (85.5, 88.0)

**Reference standard includes referral for: pure-tone average (0.5, 1, 2, 4 kHz) >25 dB OR any tone (0.5, 1, 2, 4, 6, 8 kHz) ≥30 db OR type B/C tympanometry OR digital otoscopy with pathological findings (occluding cerumen, retraction, effusion, acute otitis media, otorrhea, perforation, patent or plugged tube, external otitis, foreign body) requiring healthcare follow-up.*

†Reference standard includes type B or C tympanometry OR otoscopy findings of retraction, effusion, acute otitis media, otorrhea, perforation, presence of tympanostomy tube, or external otitis requiring healthcare follow-up.

DPOAE, distortion product otoacoustic emissions; HF, high frequency (add 6 kHz); LR+, positive likelihood ratio; LR–, negative likelihood ratio; MS, mHealth Screen (1, 2, 4 kHz); NPV, negative predictive value; PPV, positive predictive value; Tymp, tympanometry.

TABLE 4. Diagnostic accuracy statistics (with 95% confidence intervals) for children aged 3 to 6* by index tool and missing data approach, by reference standard

Reference Standard and Index Tool	N	Sensitivity	Specificity	PPV	NPV	LR+	LR–	Youden Index
High-frequency gold standard (>25dB)†								
MS								
Complete case	2541	22.4 (12.4, 32.4)	94.9 (92.6, 97.2)	45.5 (28.5, 62.4)	86.5 (83.1, 89.9)	4.4 (1.60, 7.13)	0.82 (0.71, 0.93)	17.3 (7.0, 27.5)
Ref MI, missing index as pass	2898	16.8 (11.3, 22.3)	94.6 (92.6, 96.6)	53.6 (40.7, 66.4)	75.3 (71.4, 79.1)	3.1 (1.54, 4.65)	0.88 (0.82, 0.94)	11.3 (5.4, 17.3)
Ref MI, missing index MI	2898	28.6 (21.7, 35.6)	93.8 (91.4, 96.1)	65.8 (55.3, 76.3)	75.8 (72.0, 79.7)	4.6 (2.59, 6.62)	0.76 (0.69, 0.84)	22.4 (15.2, 29.5)
HF MS								
Complete case	2541	22.4 (12.4, 32.4)	93.7 (91.2, 96.3)	40.5 (24.7, 56.4)	86.4 (82.9, 89.8)	3.6 (1.42, 5.72)	0.83 (0.72, 0.94)	16.1 (5.8, 26.4)
Ref MI, missing index as pass	2898	18.2 (12.5, 23.9)	93.1 (90.9, 95.4)	49.7 (37.7, 61.8)	75.3 (71.4, 79.2)	2.7 (1.44, 3.87)	0.88 (0.81, 0.94)	11.3 (5.2, 17.5)
Ref MI, missing index MI	2898	31.8 (24.6, 39.1)	91.9 (89.2, 94.5)	62.1 (52.1, 72.1)	76.3 (72.4, 80.1)	3.9 (2.41, 5.43)	0.74 (0.66, 0.82)	23.7 (16.2, 31.2)
DPOAE								
Complete case	2560	55.7 (44.1, 67.4)	94.6 (92.3, 96.9)	66.1 (54.0, 78.2)	91.9 (89.1, 94.6)	10.3 (5.43, 15.24)	0.47 (0.34, 0.59)	50.3 (38.5, 62.2)
Ref MI, missing index as pass	2898	47.4 (39.8, 55.0)	93.7 (91.6, 95.9)	73.9 (65.6, 82.1)	82.7 (79.0, 86.3)	7.6 (4.63, 10.56)	0.56 (0.48, 0.64)	41.2 (33.2, 49.1)
Ref MI, missing index MI	2898	45.7 (38.6, 52.9)	93.8 (91.6, 96.1)	75.6 (67.7, 83.6)	80.5 (76.8, 84.2)	7.4 (4.48, 10.37)	0.58 (0.50, 0.66)	39.6 (32.0, 47.1)
MS + Tympanometry								
Complete case	2525	66.7 (55.3, 78.0)	94.0 (91.5, 96.5)	67.7 (56.3, 79.1)	93.7 (91.2, 96.3)	11.1 (6.11, 16.05)	0.35 (0.23, 0.48)	60.6 (49.0, 72.3)
Ref MI, missing index as pass	2898	58.0 (49.7, 66.4)	93.2 (90.9, 95.4)	76.0 (68.8, 83.2)	85.6 (81.8, 89.4)	8.5 (5.27, 11.72)	0.45 (0.36, 0.54)	51.2 (42.3, 60.1)
Ref MI, missing index MI	2898	62.9 (55.4, 70.4)	92.4 (89.7, 95.1)	77.7 (70.7, 84.6)	85.6 (82.1, 89.1)	8.3 (5.35, 11.31)	0.40 (0.32, 0.48)	55.3 (47.7, 63.0)
HF MS + Tympanometry								
Complete case	2525	66.7 (55.3, 78.0)	92.8 (90.1, 95.5)	63.8 (52.4, 75.1)	93.6 (91.1, 96.2)	9.3 (5.45, 13.16)	0.36 (0.24, 0.48)	59.5 (47.8, 71.2)
Ref MI, missing index as pass	2898	59.3 (51.4, 67.3)	91.9 (89.5, 94.3)	73.1 (66.0, 80.3)	85.8 (82.1, 89.5)	7.3 (4.89, 9.72)	0.44 (0.35, 0.53)	51.2 (42.8, 59.6)
Ref MI, missing index MI	2898	65.3 (57.8, 72.8)	90.7 (87.9, 93.6)	74.7 (67.8, 81.5)	86.2 (82.6, 89.8)	7.1 (4.80, 9.31)	0.38 (0.30, 0.47)	56.0 (48.2, 63.8)
DPOAE + Tympanometry								
Complete case	2542	78.3 (68.5, 88.0)	93.8 (91.3, 96.2)	70.1 (59.9, 80.4)	95.8 (93.8, 97.9)	12.6 (7.35, 17.76)	0.23 (0.13, 0.34)	72.0 (62.0, 82.1)
Ref MI, missing index as pass	2898	66.8 (59.0, 74.6)	92.5 (90.1, 94.8)	76.8 (70.1, 83.5)	88.2 (84.7, 91.6)	8.9 (5.86, 11.87)	0.36 (0.27, 0.44)	59.2 (51.0, 67.5)
Ref MI, missing index MI	2898	64.6 (57.5, 71.8)	92.6 (90.1, 95.0)	78.4 (71.9, 85.0)	86.2 (82.7, 89.7)	8.7 (5.73, 11.67)	0.38 (0.30, 0.46)	57.2 (49.7, 64.7)

(Continued)

TABLE 4. (Continued)

Reference Standard and Index Tool	N	Sensitivity	Specificity	PPV	NPV	LR+	LR–	Youden Index
Middle ear disease‡								
MS								
Complete case	2733	26.1 (17.1, 35.1)	92.6 (90.2, 94.9)	40.7 (28.1, 53.2)	86.5 (83.5, 89.5)	3.5 (1.86, 5.16)	0.80 (0.70, 0.90)	18.7 (9.4, 27.9)
Ref MI, missing index as pass	2898	19.5 (12.6, 26.4)	94.0 (92.0, 95.9)	42.0 (29.4, 54.7)	83.9 (81.0, 86.7)	3.2 (1.67, 4.79)	0.86 (0.78, 0.93)	13.4 (6.2, 20.6)
Ref MI, missing index MI	2898	29.7 (21.5, 37.8)	91.0 (88.2, 93.7)	42.6 (31.5, 53.7)	85.1 (82.3, 88.0)	3.3 (1.96, 4.63)	0.77 (0.68, 0.87)	20.6 (12.1, 29.1)
HF MS								
Complete case	2733	26.1 (17.1, 35.1)	90.4 (87.8, 93.1)	34.8 (23.5, 46.0)	86.2 (83.2, 89.3)	2.7 (1.52, 3.94)	0.82 (0.72, 0.92)	16.5 (7.2, 25.9)
Ref MI, missing index as pass	2898	19.5 (12.6, 26.4)	92.2 (90.0, 94.4)	35.9 (24.6, 47.3)	83.6 (80.7, 86.5)	2.5 (1.36, 3.64)	0.87 (0.80, 0.95)	11.7 (4.4, 19.0)
Ref MI, missing index MI	2898	31.1 (22.7, 39.4)	88.5 (85.4, 91.5)	37.8 (27.9, 47.8)	85.0 (82.1, 87.9)	2.7 (1.70, 3.70)	0.78 (0.68, 0.88)	19.5 (10.7, 28.3)
DPOAE								
Complete case	2840	57.1 (48.3, 66.0)	91.3 (88.9, 93.6)	58.6 (49.7, 67.6)	90.8 (88.4, 93.2)	6.5 (4.51, 8.59)	0.47 (0.37, 0.57)	48.4 (39.2, 57.6)
Ref MI, missing index as pass	2898	57.7 (49.0, 66.3)	91.6 (89.3, 93.9)	60.7 (51.6, 69.7)	90.6 (88.2, 93.0)	6.9 (4.67, 9.07)	0.46 (0.37, 0.56)	49.3 (40.2, 58.3)
Ref MI, missing index MI	2898	58.8 (50.2, 67.5)	91.4 (89.0, 93.8)	60.7 (51.9, 69.5)	90.8 (88.4, 93.2)	6.8 (4.71, 8.98)	0.45 (0.35, 0.55)	50.2 (41.2, 59.2)
MS + Tympanometry								
Complete case	2711	82.0 (74.0, 90.0)	91.0 (88.5, 93.6)	63.5 (54.7, 72.3)	96.4 (94.7, 98.1)	9.2 (6.37, 11.95)	0.20 (0.11, 0.29)	73.1 (64.7, 81.5)
Ref MI, missing index as pass	2898	79.3 (72.2, 86.4)	92.4 (90.2, 94.6)	70.1 (62.3, 77.8)	95.2 (93.4, 97.0)	10.4 (7.21, 13.65)	0.22 (0.15, 0.30)	71.7 (64.2, 79.1)
Ref MI, missing index MI	2898	84.0 (77.1, 90.8)	89.7 (86.6, 92.7)	64.7 (56.4, 73.1)	96.1 (94.4, 97.9)	8.2 (5.69, 10.62)	0.18 (0.10, 0.26)	73.6 (66.1, 81.1)
HF MS+ Tympanometry								
Complete case	2711	82.0 (74.0, 90.0)	89.1 (86.3, 91.9)	58.9 (50.2, 67.5)	96.3 (94.5, 98.1)	7.5 (5.46, 9.63)	0.20 (0.11, 0.29)	71.1 (62.7, 79.6)
Ref MI, missing index as pass	2898	79.3 (72.2, 86.4)	90.8 (88.4, 93.2)	65.9 (58.2, 73.7)	95.1 (93.3, 96.9)	8.6 (6.21, 11.04)	0.23 (0.15, 0.31)	70.1 (62.5, 77.6)
Ref MI, missing index MI	2898	84.2 (77.6, 90.8)	87.4 (84.2, 90.6)	60.1 (52.1, 68.1)	96.1 (94.3, 97.8)	6.7 (4.91, 8.47)	0.18 (0.10, 0.26)	71.6 (64.2, 79.0)
DPOAE + Tympanometry								
Complete case	2817	87.9 (82.0, 93.9)	90.1 (87.7, 92.6)	65.4 (57.9, 72.9)	97.2 (95.8, 98.7)	8.9 (6.58, 11.26)	0.13 (0.07, 0.20)	78.1 (71.6, 84.5)
Ref MI, missing index as pass	2898	85.7 (79.6, 91.8)	90.3 (87.8, 92.8)	66.5 (59.1, 74.0)	96.6 (95.0, 98.1)	8.9 (6.48, 11.22)	0.16 (0.09, 0.23)	76.0 (69.4, 82.7)
Ref MI, missing index MI	2898	88.1 (82.2, 94.0)	90.1 (87.6, 92.6)	66.7 (59.4, 74.0)	97.1 (95.6, 98.6)	8.9 (6.55, 11.21)	0.13 (0.07, 0.20)	78.2 (71.8, 84.6)

**Age-specific estimates were produced using regression on the full sample (all ages) with interaction effects by age, thus sample sizes reflect number of observations used across all ages rather than number of children age 3 to 6 in the sample.*

†Reference standard includes referral for: pure-tone average (0.5, 1, 2, 4 kHz) >25 dB OR any tone (0.5, 1, 2, 4, 6, 8 kHz) ≥30 dB OR type B/C tympanometry OR digital otoscopy with pathological findings (occluding cerumen, retraction, effusion, acute otitis media, otorrhea, perforation, patent or plugged tube, external otitis, foreign body) requiring healthcare follow-up.

‡Reference standard includes type B or C tympanometry OR otoscopy findings of retraction, effusion, acute otitis media, otorrhea, perforation, presence of tympanostomy tube, or external otitis requiring healthcare follow-up.

DPOAE, distortion product otoacoustic emissions; HF, high frequency (add 6 kHz); LR+, positive likelihood ratio; LR–, negative likelihood ratio; MI, multiple imputation; MS, mHealth Screen (1, 2, 4 kHz); PPV, positive predictive value; Tymp, Tympanometry; NPV, negative predictive value.

TABLE 5. Diagnostic accuracy statistics (with 95% confidence intervals) by index tool and reference standard for children age 7 years of age and older*, complete case analysis

Tool	N	Sensitivity	Specificity	PPV	NPV	LR+	LR−	Youden Index
High-frequency gold standard (>25 dB)†								
MS	2541	42.9 (38.4, 47.4)	94.9 (93.9, 96.0)	70.6 (65.4, 75.9)	85.4 (83.7, 87.0)	8.4 (6.47, 10.40)	0.60 (0.55, 0.65)	37.8 (33.2, 42.4)
HF MS	2541	52.9 (48.4, 57.4)	93.7 (92.5, 94.9)	70.5 (65.8, 75.3)	87.5 (85.9, 89.0)	8.4 (6.68, 10.12)	0.50 (0.45, 0.55)	46.6 (41.9, 51.2)
DPOAE	2560	58.0 (53.5, 62.4)	90.6 (89.2, 92.0)	63.8 (59.2, 68.3)	88.3 (86.8, 89.8)	6.2 (5.12, 7.20)	0.46 (0.41, 0.51)	48.6 (43.9, 53.2)
MS + Tymp	2525	59.3 (54.8, 63.7)	94.0 (92.8, 95.1)	73.5 (69.1, 78.0)	89.1 (87.6, 90.6)	9.9 (7.83, 11.88)	0.43 (0.39, 0.48)	53.3 (48.6, 57.9)
HF MS + Tymp	2525	67.0 (62.7, 71.3)	92.8 (91.5, 94.0)	72.3 (68.1, 76.6)	90.9 (89.5, 92.3)	9.3 (7.56, 10.98)	0.36 (0.31, 0.40)	59.8 (55.3, 64.3)
DPOAE + Tymp	2542	66.6 (62.3, 70.9)	89.8 (88.4, 91.3)	64.9 (60.6, 69.2)	90.5 (89.1, 91.9)	6.5 (5.51, 7.57)	0.37 (0.32, 0.42)	56.4 (51.9, 60.9)
High-frequency gold standard (≥20 dB)‡								
MS	2541	38.1 (34.3, 41.9)	97.0 (96.2, 97.9)	84.6 (80.4, 88.8)	78.6 (76.7, 80.5)	12.9 (8.93, 16.85)	0.64 (0.60, 0.68)	35.2 (31.3, 39.0)
HF MS	2541	46.1 (42.3, 50.0)	96.0 (95.0, 97.0)	83.0 (79.1, 86.9)	80.7 (78.8, 82.5)	11.4 (8.45, 14.44)	0.56 (0.52, 0.60)	42.1 (38.1, 46.1)
DPOAE	2560	49.9 (46.0, 53.8)	92.6 (91.2, 93.9)	74.3 (70.2, 78.4)	81.1 (79.3, 83.0)	6.7 (5.41, 8.04)	0.54 (0.50, 0.58)	42.5 (38.4, 46.6)
MS + Tymp	2525	50.5 (46.6, 54.4)	96.1 (95.1, 97.1)	84.5 (80.8, 88.2)	82.1 (80.3, 83.9)	12.9 (9.51, 16.32)	0.52 (0.47, 0.56)	46.6 (42.5, 50.6)
HF MS + Tymp	2525	56.9 (53.0, 60.7)	95.0 (93.9, 96.1)	82.8 (79.2, 86.4)	83.9 (82.2, 85.7)	11.4 (8.75, 14.05)	0.45 (0.41, 0.50)	51.9 (47.8, 55.9)
DPOAE+Tymp	2542	56.4 (52.5, 60.3)	91.7 (90.3, 93.1)	74.2 (70.2, 78.1)	83.2 (81.4, 85.0)	6.8 (5.54, 8.01)	0.48 (0.43, 0.52)	48.1 (43.9, 52.2)
Middle ear disease§								
MS	2733	53.0 (46.8, 59.3)	91.4 (90.1, 92.6)	44.1 (38.5, 49.8)	93.8 (92.7, 94.9)	6.1 (5.00, 7.29)	0.51 (0.45, 0.58)	44.4 (38.1, 50.8)
HF MS	2733	59.1 (53.0, 65.2)	88.6 (87.1, 90.0)	39.9 (34.9, 44.9)	94.4 (93.3, 95.5)	5.2 (4.33, 6.00)	0.46 (0.39, 0.53)	47.7 (41.4, 54.0)
DPOAE	2840	72.6 (67.0, 78.1)	86.5 (85.0, 88.0)	40.9 (36.3, 45.5)	96.1 (95.2, 97.0)	5.4 (4.64, 6.10)	0.32 (0.25, 0.38)	59.1 (53.3, 64.8)
MS + Tymp	2711	82.8 (78.0, 87.6)	90.3 (89.0, 91.6)	51.4 (46.4, 56.4)	97.7 (97.0, 98.4)	8.5 (7.26, 9.79)	0.19 (0.14, 0.24)	73.1 (68.1, 78.0)
HF MS + Tymp	2711	85.3 (80.8, 89.8)	87.6 (86.1, 89.0)	46.0 (41.4, 50.7)	98.0 (97.3, 98.6)	6.9 (5.97, 7.76)	0.17 (0.12, 0.22)	72.9 (68.1, 77.6)
DPOAE + Tymp	2817	88.3 (84.2, 92.4)	85.6 (84.0, 87.2)	43.3 (38.9, 47.7)	98.3 (97.7, 98.9)	6.1 (5.40, 6.85)	0.14 (0.09, 0.18)	73.9 (69.5, 78.2)

**Age-specific estimates were produced using regression on the full sample (all ages) with interaction effects by age, thus sample sizes reflect number of observations used across all ages rather than number of children age 7 years of age and older in the sample.*

†Reference standard includes referral for: pure-tone average (0.5, 1,2, 4 kHz) >25 dB OR any tone (0.5, 1, 2, 4, 6, 8 kHz) ≥30 db OR type B/C tympanometry OR digital otoscopy with pathological findings (occluding cerumen, retraction, effusion, acute otitis media, otorrhea, perforation, patent or plugged tube, external otitis, foreign body) requiring healthcare follow-up.

‡Reference standard includes referral for: pure-tone average (0.5, 1,2, 4 kHz) >20 dB OR any tone (0.5, 1, 2, 4, 6, 8 kHz) ≥30 db OR type B/C tympanometry OR digital otoscopy with pathological findings (occluding cerumen, retraction, effusion, acute otitis media, otorrhea, perforation, patent or plugged tube, external otitis, foreign body) requiring healthcare follow-up.

§Reference standard includes type B or C tympanometry OR otoscopy findings of retraction, effusion, acute otitis media, otorrhea, perforation, presence of tympanostomy tube, or external otitis requiring healthcare follow-up.

DPOAE, distortion product otoacoustic emissions; HF, high frequency (add 6 kHz); LR+, positive likelihood ratio; LR−, negative likelihood ratio; MS, mHealth Screen (1, 2, 4 kHz); PPV, positive predictive value; Tymp, Tympanometry; NPV, negative predictive value.

imputation was conducted to understand how each of these tools might perform, particularly in the younger age group (3 to 6 years), where behavioral responses could not be obtained, and results were not missing at random. However, the ability of multiple imputation to produce unbiased estimates relies on the availability of enough variables that predict missingness and missing data values themselves (Sterne et al. 2009). In addition, the use of a comprehensive reference standard, while essential for ensuring children with middle ear disease are identified, reduces our ability to compare results of diagnostic accuracy across studies.

Strengths

To our knowledge, this is the first study to directly evaluate both behavioral pure-tone screening and objective OAE screening tools, as well as the inclusion of tympanometry in children across all school-ages (3 to 18+ years). Furthermore, a large dataset was used to evaluate these various screening protocols. The reference standard or benchmark assessment was completed on all children, in contrast with many studies that include a benchmark on only a portion of the sample. Our reference standard also included otoscopy and tympanometry to prioritize detection of both hearing loss and/or middle ear disease in this population.

Clinical Implications

There are substantial public health implications of this study. The addition of tympanometry to the mHealth pure-tone screening for older children and DPOAE screening for younger children (3 to 6 years of age) significantly improved the sensitivity of the tools without decreasing specificity. Tympanometry helped to more accurately identify children in need of additional testing and decreased the number of children missed with mild hearing loss and middle ear disease. Our findings support the addition of tympanometry as a required element in screening programs for rural and underserved areas globally, where ear infections are common.

Hearing screening protocols should differ for younger (3 to 6 years) and older school children (aged 7 and older). Our findings indicate that DPOAE plus tympanometry is the preferred screening protocol in children 3 to 6 years of age (preschool to 1st grade). DPOAE plus tympanometry is easier to obtain in the youngest cohort and is more accurate when compared with other screening tools that involve behavioral responses. Our findings suggest that the mHealth pure-tone screen with high frequency and tympanometry is the preferred screening protocol in children ages 7 and older. Depending on the needs of a screening program, however, DPOAE could be used interchangeably with the mHealth pure-tone screen when combined with tympanometry.

Standardized use of the most appropriate protocol for a given population is a critical component of any screening program, yet there continues to be a lack of consensus in the literature on which screening tools are the most accurate (Prieve et al. 2015). Results from our study should inform guidelines for screening and are generalizable to rural populations globally where infection-related hearing loss is most prevalent.

ACKNOWLEDGMENTS

We would like to acknowledge the participating Bering Strait communities, especially community members who facilitated this project through community engagement, study participation, and school and clinic involvement. We also appreciate the contributions of the Alaska Stakeholder Team, who provided countless hours of guidance and insight.

This study was supported by the Patient Centered Outcomes Research Institute (PCORI-AD-1602-34571, AD-1602-34571-S1). We would also like to acknowledge the T32 training grant (DC005360), which supported Kelli Hicks.

Drs. Susan Emmett and Samantha Kleindienst Robler (Principal Investigators) had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: S.K.R., A.P., P.H., J.R.E., J.J.G., A.L., N-Y.W., S.D.E.; Acquisition of data: S.K.R., S.M.I., C.D.J., P.H., A.A.R., S.D.E.; Analysis and interpretation of data: S.K.R, A.P., J.R.E., S.D.E.; Literature search: S.K.R, A.A.R., C.D.J., S.D.E.; Drafting of the manuscript: S.K.R., A.P., C.D.J., S.D.E.; Critical revision of the manuscript for important intellectual content: S.K.R., A.P., J.R.E., J.J.G., S.D.E.; Statistical analysis: A.P., J.R.E.; Obtained funding: S.D.E., P.H., S.K.R.; Study supervision: S.D.E., S.K.R.

ClinicalTrials.gov Registry No.: NCT03309553, NCT03662256.

De-identified data that support the findings of this study are available upon reasonable written request, with necessary approvals and execution of a formal data use agreement.

The authors have no conflicts of interest to disclose.

Address for correspondence: Samantha Kleindienst Robler, Department of Otolaryngology-Head & Neck Surgery, University of Arkansas for Medical Sciences, 4301 W. Markham Street, #543, Little Rock, AR 72205, USA. E-mail: skrobler@uams.edu

Received April 19, 2022; accepted November 28, 2022

REFERENCES

Alaska Statutes. (2019). http://www.akleg.gov/basis/statutes.asp#14.30.127

Anderson, K. L., Bright, K., Eichwald, J., III, J. H., Hanks, W. D., Ingrao, B., Mauceri, P., Miller, K., Tanner, G. (2011). *American Academy of Audiology Childhood Hearing Screening Guidelines.* https://www.audiology.org/wp-content/uploads/2021/05/ChildhoodScreeningGuidelines.pdf_5399751c9ec216.42663963.pdf.

Bess, F. H., Dodd-Murphy, J., Parker, R. A. (1998). Children with minimal sensorineural hearing loss: Prevalence, educational performance, and functional status. *Ear Hear*, *19*, 339–354

Bossuyt, P. M., Cohen, J. F., Gatsonis, C. A., Korevaar, D. A., group, S. (2016). STARD 2015: Updated reporting guidelines for all diagnostic accuracy studies. *Ann Transl Med*, *4*, 85.

Carroll, M., Cullen, T., Ferguson, S., Hogge, N., Horton, M., Kokesh, J. (2011). Innovation in Indian healthcare: Using health information technology to achieve health equity for American Indian and Alaska Native populations. *Perspectives in health information management/AHIMA, American Health Information Management Association*, *8*(Winter).

Curns, A. T., Holman, R. C., Shay, D. K., Cheek, J. E., Kaufman, S. F., Singleton, R. J., Anderson, L. J. (2002). Outpatient and hospital visits associated with otitis media among American Indian and Alaska native children younger than 5 years. *Pediatrics*, *109*, e41.

Driscoll, C., Kei, J., McPherson, B. (2001). Outcomes of transient evoked otoacoustic emission testing in 6-year-old school children: A comparison with pure tone screening and tympanometry. *Int J Pediatr Otorhinolaryngol*, *57*, 67–76.

Emmett, S. D., & Francis, H. W. (2015). The socioeconomic impact of hearing loss in U.S. adults. *Otol Neurotol*, *36*, 545–550.

Emmett, S. D., Platt, A., Turner, E. L., Gallo, J. J., Labrique, A. B., Inglis, S. M., Jenson, C. D., Parnell, H. E., Wang, N. Y., Hicks, K. L., Egger, J. R., Halpin, P. F., Yong, M., Ballreich, J., Robler, S. K. (2022). Mobile health school screening and telemedicine referral to improve access to specialty care in rural Alaska: A cluster-randomised controlled trial. *Lancet Glob Health*, *10*, e1023–e1033.

Emmett, S. D., Robler, S. K., Gallo, J. J., Wang, N. Y., Labrique, A., Hofstetter, P. (2019a). Hearing Norton Sound: Mixed methods protocol of a community randomised trial to address childhood hearing loss in rural Alaska. *BMJ Open*, *9*, e023081.

Emmett, S. D., Robler, S. K., Wang, N. Y., Labrique, A., Gallo, J. J., Hofstetter, P. (2019b). Hearing Norton Sound: A community randomised

trial protocol to address childhood hearing loss in rural Alaska. *BMJ Open*, *9*, e023078.

FitzZaland, R. E., & Zink, G. D. (1984). A comparative study of hearing screening procedures. *Ear Hear*, *5*, 205–210.

Flanary, V. A., Flanary, C. J., Colombo, J., Kloss, D. (1999). Mass hearing screening in kindergarten students. *Int J Pediatr Otorhinolaryngol*, *50*, 93–98.

Genders, T. S., Spronk, S., Stijnen, T., Steyerberg, E. W., Lesaffre, E., Hunink, M. G. (2012). Methods for calculating sensitivity and specificity of clustered data: A tutorial. *Radiology*, *265*, 910–916.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., Peters, J. (1997). From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss. *Ear Hear*, *18*, 440–455.

Haile, L. M., Kamenov, K., Briant, P. S., Orji, A. U., Steinmetz, J. D., Abdoli, A., Abdollahi, M., Abu-Gharbieh, E., Afshin, A., Ahmed, H., Ahmed Rashid, T., Akalu, Y., Alahdab, F., Alanezi, F. M., Alanzi, T. M., Al Hamad, H., Ali, L., Alipour, V., Al-Raddadi, R. M., Chadha, S. (2021). Hearing loss prevalence and years lived with disability, 1990–2019: Findings from the Global Burden of Disease Study 2019. *Lancet*, *397*, 996–1009.

Hofstetter, P. J., Kokesh, J., Ferguson, A. S., Hood, L. J. (2010). The impact of telehealth on wait time for ENT specialty care. *Telemed e-Health*, *16*, 551–556.

Järvelin, M. R., Mäki–torkko, E., Sorri, M. J., Rantakallio, P. T. (1997). Effect of hearing impairment on educational outcomes and employment up to the age of 25 years in northern Finland. *Br J Audiol*, *31*, 165–175.

Jung, D., & Bhattacharyya, N. (2012). Association of hearing loss with decreased employment and income among adults in the United States. *Ann Otol Rhinol Laryngol*, *121*, 771–775.

Kennedy, C. R., McCann, D. C., Campbell, M. J., Law, C. M., Mullee, M., Petrou, S., Watkin, P., Worsfold, S., Yuen, H. M., Stevenson, J. (2006). Language ability after early detection of permanent childhood hearing impairment. *N Engl J Med*, *354*, 2131–2141.

Khairi Md Daud, M., Noor, R. M., Rahman, N. A., Sidek, D. S., Mohamad, A. (2010). The effect of mild hearing loss on academic performance in primary school children. *Int J Pediatr Otorhinolaryngol*, *74*, 67–70.

Kokesh, J., Ferguson, A. S., Patricoski, C. (2011). The Alaska experience using store-and-forward telemedicine for ENT care in Alaska. *Otolaryngol Clin North Am*, *44*, 1359–1374.

Leisenring, W., Pepe, M. S., Longton, G. (1997). A marginal regression modelling framework for evaluating medical diagnostic tests. *Stat Med*, *16*, 1263–1281.

Lieu, J. E., Tye-Murray, N., Fu, Q. (2012). Longitudinal study of children with unilateral hearing loss. *Laryngoscope*, *122*, 2088–2095.

Lyons, A., Kei, J., Driscoll, C. (2004). Distortion product otoacoustic emissions in children at school entry: A comparison with pure-tone screening and tympanometry results. *J Am Acad Audiol*, *15*, 702–715.

Mahomed-Asmail, F., Swanepoel, W., Eikelboom, R. H., Myburgh, H. C., Hall, J. (2016). Clinical Validity of hearScreen™ Smartphone Hearing Screening for School Children. *Ear Hear*, *37*, e11–e17.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics*, *106*, e43–e43.

Naaktgeboren, C. A., de Groot, J. A., Rutjes, A. W., Bossuyt, P. M., Reitsma, J. B., Moons, K. G. (2016). Anticipating missing reference standard data when planning diagnostic accuracy studies. *BMJ*, *352*, i402.

Olusanya, B. O., Neumann, K. J., Saunders, J. E. (2014). The global burden of disabling hearing impairment: A call to action. *Bull World Health Organ*, *92*, 367–373.

Prieve, B. A., Schooling, T., Venediktov, R., Franceschini, N. (2015). An evidence-based systematic review on the diagnostic accuracy of hearing screening instruments for preschool-and school-age children. *Am J Audiol*, *24*, 250–267.

Reed, D., Struve, S., Maynard, J. E. (1967). Otitis media and hearing deficiency among Eskimo children: A cohort study. *Am J Public Health Nations Health*, *57*, 1657–1662.

Roberts, M. E. (1976). Comparative study of pure-tone, impedance, and otoscopic hearing screening methods. A survey of native Indian children in British Columbia. *Arch Otolaryngol*, *102*, 690–694.

Robinshaw, H. M. (1995). Early intervention for hearing impairment: Differences in the timing of communicative and linguistic development. *Br J Audiol*, *29*, 315–334.

Rutter, C. M. (2000). Bootstrap estimation of diagnostic accuracy with patient-clustered data. *Acad Radiol*, *7*, 413–419.

Sabo, M. P., Winston, R., Macias, J. D. (2000). Comparison of pure tone and transient otoacoustic emissions screening in a grade school population. *Am J Otol*, *21*, 88–91.

Sekhar, D. L., Zalewski, T. R., Beiler, J. S., Czarnecki, B., Barr, A. L., King, T. S., Paul, I. M. (2016). The sensitivity of adolescent school-based hearing screens is significantly improved by adding high frequencies. *J Sch Nurs*, *32*, 416–422.

Sekhar, D. L., Zalewski, T. R., Paul, I. M. (2013). Variability of state school-based hearing screening protocols in the United States. *J Community Health*, *38*, 569–574.

Singleton, R. J., Holman, R. C., Plant, R., Yorita, K. L., Holve, S., Paisano, E. L., Cheek, J. E. (2009). Trends in otitis media and myringtomy with tube placement among American Indian/Alaska native children and the US general population of children. *Pediatr Infect Dis J*, *28*, 102–107.

Singleton, R., Seeman, S., Grinnell, M., Bulkow, L., Kokesh, J., Emmett, S., Holve, S., McCollum, J., Hennessy, T. (2018). Trends in otitis media and myringotomy with tube placement among American Indian and Alaska Native Children and the US General Population of Children After Introduction of the 13-valent Pneumococcal Conjugate Vaccine. *Pediatr Infect Dis J*, *37*, e6–e12.

Skarzynski, H., & Piotrowska, A. (2012). Screening for pre-school and school-age hearing problems: European Consensus Statement. *Int J Pediatr Otorhinolaryngol*, *76*, 120–121.

Sliwa, L., Hatzopoulos, S., Kochanek, K., Pilka, A., Senderski, A., Skarzynski, P. H. (2011). A comparison of audiometric and objective methods in hearing screening of school children. A preliminary study. *Int J Pediatr Otorhinolaryngol*, *75*, 483–488.

Smith, D. F., & Boss, E. F. (2010). Racial/ethnic and socioeconomic disparities in the prevalence and treatment of otitis media in children in the United States. *Laryngoscope*, *120*, 2306–2312.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *Bmj*, *338*, b2393.

Swanepoel, W., Maclennan-Smith, F., Hall, J. W. (2013). Diagnostic pure-tone audiometry in schools: Mobile testing without a sound-treated environment. *J Am Acad Audiol*, *24*, 992–1000.

Swanepoel, W., Myburgh, H. C., Howe, D. M., Mahomed, F., Eikelboom, R. H. (2014). Smartphone hearing screening with integrated quality control and data management. *Int J Audiol*, *53*, 841–849.

Thompson, G. P., Sladen, D. P., Borst, B. J., Still, O. L. (2015). Accuracy of a tablet audiometer for measuring behavioral hearing thresholds in a clinical population. *Otolaryngol Head Neck Surg*, *153*, 838–842.

Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., Moeller, M. P. (2015). Language outcomes in young children with mild to severe hearing loss. *Ear Hear*, *36*, 76S–91S.

Wake, M., Hughes, E. K., Poulakis, Z., Collins, C., Rickards, F. W. (2004). Outcomes of children with mild-profound congenital hearing loss at 7 to 8 years: A population study. *Ear Hear*, *25*, 1–8.

White, I. R., Royston, P., Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*, *30*, 377–399.

Whiting, P., Rutjes, A. W., Reitsma, J. B., Glas, A. S., Bossuyt, P. M., Kleijnen, J. (2004). Sources of variation and bias in studies of diagnostic accuracy: A systematic review. *Ann Intern Med*, *140*, 189–202.

World Health Organization. (2012). WHO Global Estimates on Prevalence of Hearing Loss. World Health Organization. https://www.who.int/pbd/deafness/WHO_GE_HL.pdf?ua=1.

World Health Organization. (2014). Prevention of blindness and deafness: Grades of hearing impairment. http://www.who.int/pbd/deafness/hearing_impairment_grades/en/

World Health Organization. (2020). Deafness and hearing loss. WHO Press. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

World Health Organization. (2021). *World Report on Hearing*. W.H. Organization. https://www.who.int/publications/i/item/world-report-on-hearing

Yong, M., Panth, N., McMahon, C. M., Thorne, P. R., Emmett, S. D. (2020). How the World's Children Hear: A Narrative Review of School Hearing Screening Programs Globally. *OTO open*, *4*. https://doi.org/10.1177/2473974X20923580.

Yousuf Hussein, S., Swanepoel, W., Biagio de Jager, L., Myburgh, H. C., Eikelboom, R. H., Hugo, J. (2016). Smartphone hearing screening in mHealth assisted community-based primary care. *J Telemed Telecare*, *22*, 405–412.

Yousuf Hussein, S., Swanepoel, W., Mahomed, F., Biagio de Jager, L. (2018). Community-based hearing screening for young children using an mHealth service-delivery model. *Global health action*, *11*, 1467077.