Screening 3-year-olds for language delay using selected parent-report measures: the jury is still out[1]

Thomas Klee (Commentary Author), School of Education, Communication & Language Sciences and Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK.

Evidence-Based Communication Assessment and Intervention, 1, 58-59, 2007

Q How accurate is screening, based on selected parent-report measures, in identifying language delay in 3-year-old Dutch children?

METHODS

Design: Population-based cohort study.

Setting: Six regions across the Netherlands.

Participants: Parents and their children, who ranged in age from 26-58 months (Mean = 39.0, SD = 2.0; N = 11,423) were invited to participate, of which 8,877 (78%) responded.

Description of test: A postal survey of parents was conducted at 'around the time of their child's third birthday', consisting of a set of separate screening questionnaires and a further questionnaire about their child's current or past language problems (hereafter, *Parent report*). Completion rate ranged from 90 – 98.8% across the five questionnaires. Five screening questionnaires were evaluated: (1) *General Language Screen* (GLS), containing 9 items translated to Dutch from English; (2) a portion of the *Van Wiechen classification scheme* (VW), containing 6 items; (3) *Language Screening Instrument-Parent Form* (LSI-PF),

---

[1] *Abstracted from:* van Agt, H. M. E., van der Stege, H. A., de Ridder-Sluiter, J. G., & de Koning, H. J. (2007). Detecting language problems: accuracy of five language screening instruments in preschool children. *Developmental Medicine & Child Neurology, 49,* 117-122.

For correspondence: E-mail: h.vanagt@erasmusmc.nl

Source of funding: none reported

containing 4 items; (4) *Language Screening Instrument-Child Test* (LSI-CT), containing 28 items; and, (5) a *Visual analogue scale* (VAS), asking parents to 'mark the place on the line below which you feel accurately represents your child's language development compared to his or her peers'.

Diagnostic standard: Two reference standards were employed: (1) a measure based on the *Parent Report* questions described above; and (2) a measure based on the diagnostic outcome of children who had previously received speech and language evaluations subsequent to referral (hereafter, *Specialist Report*). For the *Parent Report* standard, case status was based on parents indicating that their child had been assessed by a specialist (speech therapist, ENT or paediatrician), that a language problem was observed, and that their child 'knew too few words for his or her age; exhibited no or insufficient spontaneous speech; or had difficulty understanding what others said'. For the *Specialist Report* standard, case status was based on the specialists' judgement that the child had a 'language problem, late start of language development, or language development below standard' (de Koning, de Ridder-Sluiter, van Agt, Reep-van den Bergh, van der Stege, Korfage, Polder & van der Mass, 2004, 111).

Outcome measures: Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive yield, and overall accuracy (area under the receiver operating characteristic curve) were calculated against each reference standard, using three cut-off scores for each screen. Agreement between reference standards, internal consistency, and concurrent validity (inter-test correlations) were also calculated.

MAIN RESULTS

Prevalence of language delay in this population was reported to be 2.9% relative to the *Parent Report* reference standard and 1.5% relative to *Specialist Report*. Sensitivity ranged from 43 to 86%, and specificity 73 to 97%, across the five screening measures, calculated

using cut-off scores which yielded approximately 5%, 10% and 20% positive screens. PPV ranged from 4 to 31%, with NPV consistently around 99%. Internal consistency ranged from .67 to .72 across screening measures, while inter-test correlations ranged from .29 to .72. Agreement between the two reference standards was .41.

AUTHORS' CONCLUSIONS

The authors concluded that 'short and simple parent report instruments like the LSI-PF and the one-item VAS perform remarkably well in detecting language delays in preschool children' (p.117). Later, they qualify this: '…none of the instruments exhibited high sensitivity for cut-off values assigning small groups of screen-positive children' (p.121). They suggest that parent-based screening measures such as these may function most effectively as the first stage of a two-stage screening procedure.

Commentary

Two systematic reviews of screening instruments for speech and language delay have been published (Law, Boyle, Harris, Harkness & Nye, 2000a; Nelson, Nygren, Walker & Panoscha, 2006), both reaching similar conclusions. The more recent of these, which formed the basis for a recommendation by the US Preventive Services Task Force (USPSTF), concluded that '…the evidence is insufficient to recommend for or against routine use of brief, formal screening instruments in primary care to detect speech and language delay in children up to 5 years of age' (USPSTF, 2006, 497). This report went on to say that more research was needed to identify effective, brief, formal instruments that could be used to screen children in this age group. Although the large scale, population-based study by van Agt and colleagues is a response to this challenge, the evidence provided in their study does not alter the conclusions of these systematic reviews for a number of reasons.

Using QUADAS, an evidence-based quality assessment tool for evaluating diagnostic accuracy studies (Whiting, Rutjes, Reitsma, Bossuyt & Kleijnen, 2003), several study design

weaknesses were noted. One of the more serious was whether the two diagnostic standards were likely to correctly classify children as language delayed or not. Neither is the diagnostic accuracy of these particular reference standards known, nor could either be considered to be a universally accepted (gold) standard – criticisms that could be levelled at any screening study published in this area. Even accepting that the *Specialists' Report* reference standard was reasonably accurate, it was completed *before* rather than *after* the screening was done. Consequently, the parent respondents would not have been blind to the diagnostic outcome at the time they completed the screening questionnaires. Moreover, since the questions making up the *Parent Report* reference standard were included in the postal survey containing the screening questionnaires, parents' responses to the screening questionnaires and the questions which formed the reference standard could not have been independent. Similarly, the fact that parents were asked to complete all five screening measures at the same time meant that the screening outcomes could not have been independent of one another. How parents responded on one questionnaire probably influenced how they responded on others. The use of the *Parent Report* and *Specialists' Report* reference standards may have resulted in netting only the most severe of cases, in that 2.9% and 1.5% of children were considered to be language delayed respectively, figures that are below other reported estimates of prevalence (Law, Boyle, Harris, Harkness & Nye, 2000b). Some of these study limitations are also discussed by the authors of this report and others are outlined in a commentary by Grether (2007), published in the same issue of the journal.

The likelihood ratios for a positive screening result (LR+), computed from the data provided in Table III of the report, ranged from 3-15, while those for a negative screening result (LR-) ranged from .19-.62. The LRs were similar regardless of which reference standard was used. None of the five screening measures, at any of the reported cut-off levels,

yielded a combination of LR+ ≥ 10 and LR- ≤ .20, casting doubt on their practical value as screening measures.

The USPSTF also pointed out in its review of screening studies from 1966-2004 that 'no studies…addressed the overarching question of whether screening for speech and language delay with brief, formal instruments results in improved speech, language, and other non-speech and language outcomes' (p. 498). To their credit, this same research team recently published a trial examining the outcome at 36 months of children who had been screened for language delay at 18 and 24 months, involving essentially the same cohort of children who participated in the study reviewed here (de Koning et al., 2004).

References

de Koning, H. J., de Ridder-Sluiter, J. G., van Agt, H. M. E., Reep-van den Bergh, C. M. M., van der Stege, H. A., Korfage, I. J., Polder, J. J. & van der Mass, P. J. (2004). A cluster-randomised trial of screening for language disorders in toddlers. *Journal of Medical Screening, 11*, 109-116.

Grether, S. (2007). Commentary on 'Detecting language problems: accuracy of five language screening instruments in preschool children'. *Developmental Medicine & Child Neurology, 49*, 84.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000a). The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Developmental Medicine & Child Neurology, 42*, 190-200.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000b). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *International Journal of Language & Communication Disorders, 35*, 165-188.

Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics, 117*, 298-319.

US Preventive Services Task Force. (2006). Screening for speech and language delay in preschool children: recommendation statement. *Pediatrics, 117*, 497-501.

Whiting, P., Rutjes, A. W. S., Reitsma, J. B., Bossuyt, P. M. M., & Kleijnen, J. (2003). The development of QUADAS: a tool for the quality of assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology, 3*, 25.