Screening for speech and language delay: a systematic review of the literature

Law J, Boyle J, Harris F, Harkness A, Nye C

Authors' objectives
To assess the value of screening and intervention for speech and language delays in children up to the age of 7 years. Two additional objectives were to identify priority areas for further investigation and to provide evidence-based direction for the future provision of services.

Searching
A preliminary literature search of the Cochrane Database of Systematic Reviews, the Database of Abstracts and Reviews of Effectiveness, the Cochrane Controlled Trials Register, and the Cochrane Review Methodology Database failed to locate any reviews on this topic.

The authors then searched the electronic databases: Cumulative Index of Nursing and Allied Health; EMBASE; Educational Resources International Clearing House; Linguistics and Language Behaviour Abstracts; MEDLINE; and PsycLIT.

The authors handsearched relevant journals, compilation volumes, and bibliographies, and used Internet search engines, and databases of unpublished literature to identify additional studies. Contact was made with other professional organisation, institutions and authors, and feedback received and incorporated during the review.

The review also made use of funnel plots of standardised effect size by sample size and by study quality to examine whether there were gaps in the data set due to publication bias.

Study selection
Study designs of evaluations included in the review
For intervention: the studies had to detail the number of participants in each group and the nature, duration, span, and delivery of treatment, provide a comparison of pre- and post- intervention speech and language measures, and fulfill one of three design criteria: experimental study with randomised non-treatment controls; quasi-experimental studies (with non-random/pseudo-random or non-equivalent non-treatment control groups); or a single-subject experimental design with graphical displays or session-by-session data for individuals.

For screening tests: studies also had to included information on sample size, whether drawn from the general population, or a clinical population, or a mixed population, and had to give clear criteria for scores and report results expressed by both sensitivity and specificity.

Specific interventions included in the review
The authors state that approaches to interventions vary considerably. The interventions covered four broad areas: didactic (elicited imitation or mand modelling) approaches; naturalistic (interactionist or incidental teaching) approaches; hybrid (milieu therapy) approaches; and other approaches (non-directive therapy, auditory training, comprehension monitoring, and cognitive therapy). Specific categories used to identify the particular area of language investigated by the study were: attention control/concentration/listening skills, other prelinguistic skills, speech, expressive language, receptive language, general language, pragmatics/social use of language, cognitive abilities, and parent-child interaction. Screening interventions had to use tests designed for use in primary health care settings and/or in an educational setting by non-specialist staff for early identification, not diagnosis.

Reference standard test against which the new test was compared
The review did not include any diagnostic accuracy studies that compared the performance of the index test with a reference standard of diagnosis.

Participants included in the review
For effectiveness of intervention: children up to 7 years of age. For accuracy of screening: normal and clinical populations within the 0-7 year age range.

Studies had to be of primary speech and language delay. Studies focusing on ADD/ADHD; deafness/sensorineural loss; autism; psychiatric or EBD; Down's Syndrome; cerebral palsy; dyslexia; effects of social disadvantage; or other secondary speech and language delay were excluded.

Outcomes assessed in the review
Impact on speech/language delay. Main areas considered were articulation/phonology, expressive language, and auditory discrimination. A wide variety of scales appear to have been used to assess outcomes.

How were decisions on the relevance of primary studies made?
Final judgements about inclusion were made by two independent assessors, with discussion to resolve any disagreements.

Assessment of study quality
An overall rating of the quality of each study was made based upon study reliability and study validity, together with additional comments. In RCTs and quasi-experimental study designs, reliability scores (maximum of 19 points) ranged between 9 and 19 points, and validity scores (maximum 15 points) ranged between 8 and 14 points. Single-subject experimental designs were grouped into higher quality studies (study validity greater or equal to 10 points) or lower quality studies (study validity less than 10 points). The validity assessment was part of the data extraction form and was performed by two independent reviewers.

Data extraction
The data were extracted using data extraction forms by two independent coders. The percentage agreement rate was calculated for each of the four domains and the overall % agreement rates were: prevalence 84.8%; natural history 85.4%; intervention 89.0%; and screening 90.2%.

Coding categories for prevalence studies were: study details; criteria for language delay; sample; prevalence; and quality rating. Coding categories for natural history studies were: study details; subject characteristics; design; areas of investigation; study characteristics; outcome findings; and quality rating. Coding categories for intervention studies were: study details; subject characteristics; design; areas of intervention; study characteristics; outcome findings; and quality rating. Coding categories for screening were: study details; screening procedure; criteria; sample; reliability; validity; and quality rating.

Methods of synthesis
How were the studies combined?
In the intervention studies, standardised effect sizes (ES) for the outcomes from randomised controlled trials (RCTs) and quasi-experimental designs and the PND statistic (percentage of non-overlapping data between baseline and post-baseline phases) for single-subject experimental designs were calculated.

For screening studies, sensitivity and specificity rates and likelihood ratios (LRs) were calculated.

How were differences between studies investigated?
The review inspected the direction of outcomes from excluded studies to examine the extent to which studies excluded on grounds of design also showed positive effects of intervention.

Results of the review
For intervention: there were 10 RCTs with 151 subjects and 106 controls; 11 quasi-experimental studies with 203 subjects and 189 controls; and 26 single-subject experimental studies with 77 subjects.

For screening: there were 33 studies of concurrent validity of screening tests with 6,756 subjects given full testing and
18,022 in the screening sample; and 5 comparison studies of concurrent validity between screening tests with 829 given full screening and 1,096 in the screening sample.

In the intervention data, results from RCTs and quasi-experimental designs show positive and statistically significant effects of intervention relative to untreated controls in all areas of speech and language skills. Comparable results for direct (clinician administered) and indirect treatment were observed in the case of expressive language. In contrast, direct intervention was more effective in the case of speech, whereas indirect intervention was more effective in the case of receptive language. Data from the single-subject experimental designs were synthesised and provide confirmatory evidence for the positive effects of intervention. The data reviewed do not provide information about the long-term outcomes of intervention, nor of the likelihood of intervention reducing prevalence in a given population. It is not possible to draw conclusions about the effects of subject variables such as socio-economic status or age upon the relative value of intervention. The screening evidence, few studies compare the performance of two or more screening tests when applied to one population, nor do they compare single screening measures across different populations. In general, specificity is higher that sensitivity, suggesting that it is easier to determine who is not a case than to establish who is. Parent-focused measures appear to be as useful as specific tests of child behaviour. There remains considerable disagreement as to what proportion of the population should be considered cases.

**Authors' conclusions**
The authors state that early speech and language delay should be a concern to those involved with child health surveillance because of the problems for the individual child, because it may indicate other co-morbid conditions such as hearing loss, developmental and behavioural difficulties, and because of the implications it may have for literacy and socialisation in school. The fact that there is not sufficient evidence to merit the introduction of universal screening does not imply that speech and language delay should not be identified by less formal checks.

**CRD commentary**
This is a good systematic review. The research questions and outcome measures are clearly stated, as are the inclusion and exclusion criteria for each outcome measure. The literature review is extensive and the search strategy is reported in the review. The study details are discussed in the text of the reviews, and summarised in tables and appendices.

The pooling of the data is appropriate and the conclusions follow from the results of the review.

**Implications of the review for practice and research**
Practice: The authors state that implications for practice include paying more attention to the role of parents in identifying children with speech and language delay, and that primary care workers should be involved in eliciting parental concerns and in making appropriate observations of children’s communication behaviours. Because of the reported value of indirect approaches to intervention, there is a case for widening the range of professionals able to promote good interactive practice in parents of young children.

Research: The implications for further research focus on:

1. Examining the impact of speech and language delay.
2. Examining the medium- and long-term effects of well described models of intervention.
3. Developing a screening measure that combines data on risk factors with parental report and professional observation.
4. Developing the predictive ability of different models of early identification and intervention needs.

**Funding**
NHS R&D Health Technology Assessment (HTA) Programme, project number 94/05/02.
Bibliographic details

PubMedID
9728296

Original Paper URL
http://www.hta.ac.uk/project.asp?PjtId=953

Indexing Status
Subject indexing assigned by NLM

MeSH
Age Factors; Child; Child, Preschool; Female; Humans; Infant; Language Development Disorders /diagnosis /epidemiology; Male; Prevalence; Speech Disorders /diagnosis /epidemiology

AccessionNumber
11998009012

Date bibliographic record published
29/02/2000

Date abstract record published
29/02/2000

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.