Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature

James Law†, James Boyle‡, Frances Harris†, Avril Harkness‡ and Chad Nye§

†Department of Language and Communication Science, City University, London, UK
‡Department of Psychology, University of Strathclyde
§Department of Communication Disorders, University of Central Florida

(Received July 1998; accepted March 1999)

Abstract

The prevalence and the natural history of primary speech and language delays were two of four domains covered in a systematic review of the literature related to screening for speech and language delay carried out for the NHS in the UK. The structure and process of the full literature review is introduced and criteria for inclusion in the two domains are specified. The resulting data set gave 16 prevalence estimates generated from 21 publications and 12 natural history studies generated from 18 publications. Results are summarized for six subdivisions of primary speech and language delays: (1) speech and/or language, (2) language only, (3) speech only, (4) expression with comprehension, (5) expression only and (6) comprehension only. Combination of the data suggests that both concurrent and predictive case definition can be problematic. Prediction improves if language is taken independently of speech and if expressive and receptive language are taken together. The results are discussed in terms of the need to develop a model of prevalence based on risk of subsequent difficulties.

Keywords: primary speech and language delay, children, prevalence, natural history, epidemiology.

Introduction

The present review was commissioned and managed by the National Health Service Centre for Reviews and Dissemination at the University of York on behalf of the

*Address correspondence to: James Law, Department of Language and Communication Science, City University, Northampton Square, London EC1V OHB, UK; e-mail: J.C.Law@city.ac.uk
National Health Technology Assessment Programme of the NHS in the UK. The overall aim of the review was to establish whether there was a case for the introduction of universal population screening for speech and language delay. An epidemiological model was adopted (Law et al. 1998) and this led to the identification and review of the four key domains of the literature: prevalence, natural history, intervention and screening. Although these domains have been treated independently in the literature to date, the position taken in the review was that they are closely associated. While it may be possible to calculate prevalence at any age given a suitable measurement, it is only a meaningful construct in the context of natural history of the condition. Similarly, both screening and intervention are tied into the need for a clear concept of case definition and appropriate targeting depends on an understanding of the course of the disability. The intervention and screening literature are reviewed elsewhere (Boyle et al. in preparation, Law et al. in press).

**Prevalence**

Prevalence is conventionally expressed as the proportion or percentage of cases in a given population at a specified time. Normal rather than clinical population is the focus in the present review. Prevalence differs from, but is often confused with, incidence, which refers to the number of new cases (Mosciki 1984). Cases may move out of the pool from which prevalence is taken because of mediating factors (such as spontaneous recovery, intervention or death), but prevalence also depends upon incidence. Prevalence will rise if incidence exceeds the decrease brought about by these mediating factors, and is an important concept for three reasons: it allows planning of service delivery; it should also allow calculation at an epidemiological level of the impact of intervention, with a successful intervention being one that results in a decline in prevalence; and it should reflect current knowledge about a disorder where the boundaries lie between for example normality and abnormality and which cases respond to intervention. To date there has been no attempt to carry out a systematic synthesis of the prevalence literature although there have been some extensive narrative reviews (MacKeith and Rutter 1972, Healey et al. 1981, Silva et al. 1987, Harasty and Reed 1994).

**Natural history**

‘Natural history’ is a specific term used to describe the prognosis of a condition in the absence of intervention (Gordis 1996). It can be a difficult concept when it comes to children’s development because, in most cases, children receive some form of intervention by virtue of entering the education system. Nonetheless there probably is a distinction to be drawn between intervention of a generic nature and intervention which is targeted at the child’s specific difficulties. Natural history is important because the anticipated status of a putative case allows determination of current status. Thus, if a two cases are identified as having comparable levels of skill at time one but it is known that there is a much greater risk of persistent problems for case one than for case two, then the relative weighting attributed to case one will be higher.
Systematic review of the literature

The aims of a systematic review are to locate, appraise and synthesize evidence from studies to provide empirical answers to scientific research questions. It differs from other types of review in that it adheres to a strict scientific design to make it more comprehensive, and to minimize the chance of bias (NHS Centre for Reviews and Dissemination 1996). Such a review follows a strict procedure that involves the identification of published and unpublished data sources, the development of research questions, search strategies and inclusion criteria, the reliable coding of the data, and its synthesis ultimately leading to data-driven conclusions.

Methodology

Scope of the literature

The literature for 1967 to May 1997 was targeted. CROS AND DIALINDEX search mechanisms identified the databases most likely to identify appropriate literature. Thus, six databases were identified: Embase, Medline, ERIC, PsychINFO, CINAHL and LLBA. In addition, the System for Indexing Grey Literature in Europe (SIGLE) and the Boston Spa Conferences (British Library Databases) were searched for unpublished, or ‘grey’, literature. Grey literature is scanned to attempt to overcome publication bias. Finally, journals were hand-searched to identify relevant authors in the field. In all, 53 prevalence papers met the criteria of relevance for prevalence but of these only 21 met the full inclusion criteria. These provided 16 sets of data. Twenty-three papers met the initial criteria for relevance in the natural history domain, but of these only 18 were included in the review providing 12 data sets (see Appendices 1 and 2).

Criteria for inclusion in the review

The review dealt with children with primary speech and/or language delays. That is, it did not include children with delays secondary to other conditions such as autism, or more general developmental disabilities. It is recognized that, in the absence of universally recognized criteria for primary delays, this may be a difficult judgement to make and that there may be more similarities than differences between these groups. Nonetheless the distinction between primary and secondary delay is one that is widely recognized by both researchers and clinicians and, therefore, has face validity. Authors use many different ways to define the populations they describe. Some use discrepancies between measures of language and general abilities, but again there is no widely accepted level of discrepancy that could be imposed on the data. In addition, some design criteria were specified for each domain.

Prevalence studies

Studies were included that estimated the prevalence of speech and language delays in children aged up to 16 years in a general population. The studies needed to present data about the number of participants and the diagnostic samples, and the definition of case status had to be determined either by standardized measures of speech and/or language or to use clearly defined clinical judgement. Studies of clinic populations were excluded. In addition, a distinction has been drawn between
single-level prevalence studies, which have been based on surveys or clinical judgement, not subjected to external validation, and those studies that have been based on a pre-screen or net of the population with proportions of passes and fails sampled and then given a diagnostic assessment, either on a standardized language procedure or on a criterion referenced clinical judgement (which the authors have made some attempt to validate or define such that it could be replicable). The ‘single level designs’ have been excluded on the grounds that they would be impossible to replicate and are likely to lead to under-reporting in all but the most clear-cut medical cases (Leske 1981). Many of the largest studies have been excluded because, although the sampling was potentially of considerable interest, their single-stage approach would not have met the inclusion criteria for the review. They have also tended to report the lowest prevalence figures (Blum-Harasty and Rosenthal 1992). This two-stage criterion resulted in the exclusion of a number of studies that are commonly included in narrative reviews most notably Drillien and Drummond (1983), Fundidis et al. (1979), Hull et al. (1971), Morley (1965) and Peckham (1973). The reason for applying such a stringent criterion is that it provides a measure of control against threats to the reliability of the results and hence to the validity of the interpretation.

Natural history

Large-scale prospective same-age cohort longitudinal studies offer the best evidence of natural history outcomes. However, most of the prospective studies located by the review were those where at least a proportion of the subjects received speech and language therapy services. A statement regarding the numbers in therapy, or the amount of therapy received, led to the study being excluded from this review. ‘Therapy’ was interpreted in the context of the studies as specific advice or assessment from speech and language therapists. Where there was no such statement about therapy contact for the subjects, the study was included. In a minority of instances a study explicitly stated there was no intervention for the subjects. General advice given routinely by a health professional other than a speech and language specialist was regarded as non-specific intervention, and such studies were included. This applied, for example, to some treatment studies where control groups received only general help. Moreover, within a prospective design with some subjects receiving treatment, where data for individual untreated subjects could be retrieved, their outcome was noted. Another criteria for inclusion ensured that there was a retest interval of at least 6 months, using norm- or criterion-referenced language outcome measures. No minimum number of subjects was set.

These constraints, in particular the stringent inclusion criteria, also resulted in a number of well-known studies being excluded from the review (e.g. Sheridan and Peckham 1975, Klackenburg 1980, Paul 1993). While their designs have many exemplary features, and have achieved long-term follow up for relatively large samples, it has not proved possible to separate out possible treatment effects in their data set.

Reliability

To ensure that the same data was being extracted appropriately, two independent coders coded a subset of the papers and the percentage agreement rate was reported. Nineteen per cent of the prevalence papers and 14% of the natural history papers
were subjected to the reliability check resulting in reliabilities of 84.8 and 85.4% respectively. Differences were resolved following discussion between the coders. The majority of sources of difference between the coders arose as the result of gauging which figures to use as the denominator at any one stage in the calculation of prevalence or natural history because of differential attrition in the populations concerned.

**Results**

**Results from prevalence studies**

To represent its diversity the data in the present review were classified by the language domains measured, and by the age of the children in the sample. Table 1 shows prevalence for (1) speech and/or language, (2) language and (3) speech. Table 2 focuses on those studies that have looked at language delay in the absence

Table 1. Median prevalence estimates by type of speech and language delay and age

<table>
<thead>
<tr>
<th>Type of delay</th>
<th>Speech/language delay, language delay only and speech delay only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Speech/language delay median of estimates</td>
</tr>
<tr>
<td>2:0</td>
<td>5.0&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>3:0</td>
<td>6.9 [5.6–8]&lt;sup&gt;abc&lt;/sup&gt;</td>
</tr>
<tr>
<td>4:6</td>
<td>5.0 [–]&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>6:0</td>
<td>–</td>
</tr>
<tr>
<td>7:0</td>
<td>–</td>
</tr>
</tbody>
</table>

<sup>a</sup>Bax et al. (1980, 1983); <sup>b</sup>Beitchman et al. (1986); <sup>c</sup>Burden et al. (1996); <sup>d</sup>Dudley and Delage (1980); <sup>e</sup>Randall et al. (1974); <sup>f</sup>Rescorla et al. (1993); <sup>g</sup> Silva et al. (1983); <sup>h</sup> Stevenson and Richman (1976), Richman et al. (1982); <sup>i</sup>Tomblin et al. (1997); <sup>j</sup>Tuomi and Ivanoff (1977); <sup>k</sup>Wong (1992).

<sup>†</sup>Beitchman et al. (1986) is the only study to include prevalence estimates for both speech and language or speech or language. This highlights the difficulty in synthesizing data in this area because both the level of difficulty and the extent to which the categories can reasonably be teased apart from the data is unclear.

Table 2. Median prevalence estimates by type of language delay and age

<table>
<thead>
<tr>
<th>Type of delay</th>
<th>Expressive and receptive language delay, expressive delay only and receptive delay only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Expressive and receptive language delay median percent (range)</td>
</tr>
<tr>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>3.01&lt;sup&gt;kb&lt;/sup&gt; [2.63–3.4]</td>
</tr>
<tr>
<td>5</td>
<td>2.14&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>7</td>
<td>2.02&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Silva et al. (1983); <sup>b</sup>Wong (1992); <sup>c</sup>Rescorla et al. (1993); <sup>d</sup>Stevenson and Richman (1976).
of speech delay and separates those language delays out into expressive/receptive delays, isolated expressive delays and isolated receptive delays. This group represents a subgroup of those quoted in table 1. Table 3 displays studies that report prevalence across different age groups and which are impossible to combine in a meaningful way. In a number of studies more than one data set is provided for a given age band. To avoid over-representing such studies in the total data set only a single (median) figure was included for each study.

_Speech and/or language delay_

The largest single group and, thus, the group most likely to be picked up in the process of early identification (the original aim of the systematic review) is that with combined speech/language difficulties, where researchers have included the possibility of either speech or language being implicated. This corresponds to column 1 of table 1, and the median for this column is 5.95%. However, although the median was adopted as the most appropriate measure of central tendency in view of the variability observed in the data, considerable caution needs to be taken in extrapolating to produce single composite prevalence estimates especially given the relatively small number of studies involved.

Taking speech and language delay as a single construct, the majority of studies took diagnostic assessment score cut-offs between $-2$ and $-1.5$ standard deviations (SD) below the mean for the standardization sample on a single measure, which automatically gives prevalence rates of between 2.28 and 6.68%. In most cases the range of findings is slightly wider than would be anticipated on this basis, ranging from 1.35 to 8% (table 1, column 1; table 3). Where more liberal cut points are adopted higher prevalence figures result and this is compounded when failure on any one of a number of measures is used as the criteria. For example, one of the highest figures came from a large population study that used multiple measures (Beitchman et al. 1986). The figure for speech or language was much higher than most other estimates (19%) being an inclusive estimate for speech or language delay and using a cut-off of $-1$ SD below the mean on a series of different speech and language measures. By contrast, a scale of clinical difficulty adopted by Paul et al. (1992) (table 3) reported much lower figures (1.35% for mild, moderate and severe cases, 0.65 for moderate and severe or what they term ‘serious’ cases only). These figures derive from a study in which case status depended upon a single objectified criteria for clinical judgement rather than performance on standardized assessments of speech and language performance. Such conservative criteria may reflect the fact that this study was carried out in Jamaica, a country where services, with fewer

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Study</th>
<th>Type of delay</th>
<th>Prevalence estimate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:0–9.0</td>
<td>Paul et al. (1992), Thorburn et al. (1991)</td>
<td>Speech and Language</td>
<td>1.35</td>
</tr>
<tr>
<td>6.0–12</td>
<td>Harasty and Reid (1994)</td>
<td>Speech and Language</td>
<td>8.0</td>
</tr>
<tr>
<td>3:0–5.0</td>
<td>Stewart et al. (1986)</td>
<td>Speech</td>
<td>1.5</td>
</tr>
<tr>
<td>5.0–7.0</td>
<td>Kirkpatrick and Ward (1984)</td>
<td>Speech</td>
<td>4.6</td>
</tr>
<tr>
<td>6.0–12</td>
<td>Harasty and Reid (1994)</td>
<td>Speech</td>
<td>12.6</td>
</tr>
<tr>
<td>12–14</td>
<td>Warr Leeper et al. (1979)</td>
<td>Speech</td>
<td>7.3</td>
</tr>
</tbody>
</table>
Prevalence and natural history of primary speech and language delay

resources available, would be likely to apply a more stringent criterion as to who should receive services.

Language delay only

Language delay in the absence of speech delay is the group most commonly identified in the studies included in the present review. Some studies have reported children with either expressive or receptive delays as one group (table 1, column 2; table 2, column 1). The variability here is also wide, giving a range of 2.02–19%. In one case a cut-off, which would anticipate a higher prevalence estimate ($-1.25$ SD), resulted in a convergent prevalence estimate of 7.4% because children had to receive two low test scores to achieve case status (Tomblin et al. 1997). The figures for expressive with receptive language delays combined show little variation within a narrow band, reflecting the conservative cut-off scores adopted in these studies and perhaps a greater confidence in the most appropriate level at which to set case status. It is of note that this most recent population study suggests a higher level of prevalence for children with language problems based on a discrepancy criterion rather than for the broader category of children with primary delays captured by most of the other studies. It is probable that the rather more liberal cut point used in this study explains the variation. But it would be wrong to over emphasize the difference. The important issue is whether the more inclusive definition of language delay results in more cases whose difficulties would not otherwise resolve being provided with appropriate intervention.

Most notable is the sharp drop between 2 and 3 years, the high early figure created by the measure of output vocabulary adopted by Rescorla et al. (1993). In the main the estimates use either single measures of language which amalgamate different components (comprehension, expression, vocabulary, verbal memory, etc.). Sometimes cases are identified by failure on any one of a number of subtests (Stevenson and Richman 1976, Dudley and Delage 1980). In other cases composite measures made up of a number of different tests are used (Tomblin et al. 1997). The evidence seems to suggest that it does not make much difference which approach is adopted. It is noteworthy that the level of prevalence rises at the 5-year mark. In part this is the function of the 10.4% figure reported by Silva et al. (1983), which sums expressive, receptive and expressive receptive cases. These are separated out in table 2, column 1. It is also possible that there is a change in expectations at this point in the child’s development, a closer approximation to the adult norm is now expected and as the expectations increase so do the number of potential cases of delay.

In a study of primary speech and language delay it is obviously important to establish what is known about the prevalence of specific language impairment (SLI), that is language impairments that are identified by means of the discrepancy between language and IQ measures. Only two studies have achieved this with very disparate results. Stevenson and Richman (1976), who obtained a prevalence of 0.6% for SLI, and Tomblin et al. (1997), who obtained 7.4% for a group of SLI children. This difference is likely to stem from the difference in discrepancy criterion adopted and the age of the children in the study. Thus, Tomblin et al. specified that IQ had to be within the normal range with language level below $-1.25$ SD on their composite measure at 5 years, while Stevenson and Richman specified that expressive language had to be less than two-thirds of mental age, which in turn had to be two-thirds of chronological age at 3 years of age.
Speech delay only

The figures for speech (not language) delay are again highly variable, ranging from 2.3 to 24.6% (table 1, column 3; table 3). Kirkpatrick and Ward (1984) identified 4.6% using a $-2$ SD cut-off. Beitchman et al. (1986) set a $-2$ SD cut-off on particular subtests to reach their speech prevalence rate of 6.4%. In contrast, Tuomi and Ivanoff (1977) adopted a criteria of 1 year behind chronological age and reported 24.6% in kindergarten children, although this fell to 16.5% by Grade 1. These figures are relatively high and probably reflect the relatively liberal definition of what constitutes a case, including include children with poor ‘stimulability’, multiple and consistent errors and those with a lateral lisp. It would appear that the childrens’ speech is reported in absolute rather than developmental terms. Significantly prevalence studies of speech delay in the pre-school years did not fulfil the inclusion criteria for the review. This probably reflects a recognition of the difficulty in establishing case status during the very early years.

Language only—expressive and receptive delay

Two studies have reported data that allow combination of expressive and receptive language scales (Silva et al. 1983, Wong et al. 1992). It might be anticipated that these data should coincide with those from table 1, column 2. In fact, this method of classifying the children suggests a rather more conservative estimate of prevalence as between 2 and 3%, which remains constant across the age range. To classify children in this way it is necessary to use standardized procedures that tease apart expressive and receptive skills. But inevitably the precision using measures in this way effectively exacerbates the circularity associated with the use of this type of metric, a point discussed further below.

Language only—expressive delay

Three authors report figures for delays only in expressive language skills. Two studies use only an expressive language measure (Stevenson and Richman 1976, Rescorla et al. 1993), while Silva et al.’s criteria (1983) seek to identify expressive delay in the absence of delay in receptive language skills. The figures reported by Stevenson and Richman, and Silva range from 2.34 to 4.27% over the age range 3–7 years. The highest figure comes from Rescorla et al. (1993), with 8, 16 and 19% according to the screen cut-off. These figures were computed from a single-vocabulary checklist rather than a diagnostic test performance, which was the approach adopted in the other studies. That is, the range of reported prevalence reflects the initial screening stage of this study. Alternative prevalence figures can be computed reflecting the reference measure, giving 9.8 and 13%. However, the age range for Rescorla et al.’s study was at least 1 year below that of the other studies, and it may be that the range of expressive vocabulary development is especially wide in the slightly lower age range.

Language only—receptive delay

The figures for receptive delay are again tightly grouped ranging from 2.63 to 3.95%. All the figures come from the Dunedin study (Silva et al. 1983). Silva et al.’s figures here could include children who are actually expressive-receptive language
delayed, but who just achieved a pass on the expressive language measure while failing the receptive test. The figures may, therefore, be an instance of psychometric convention for cut-offs confounding the clinical impression.

Relying on standardized assessments to establish case status makes it difficult to judge whether prevalence decreases with age. A given cut-off on such a measure will result in the same percentage of the population being identified at any given time. The data presented here suggest that prevalence does not decrease over time. Bax et al. (1983) indicate fairly stable figures over the 2–4.5-year age range for their ‘definitely abnormal’ group. It seems likely that the majority of studies reviewed above take this latter relatively stable group as their subject pool. This could suggest that similar prevalence figures across a time range reflect the same group of children at different points in time. In fact, the evidence from Silva et al. (1983) is that individual children may move in and out of the disordered group. This may be true or it may be a function of the lack of stability of the metrics adopted. Predictability of problems depends upon the presenting symptoms at any one point, children with expressive and receptive delays being much more likely to have persistent problems than children with expressive delays alone. By contrast, Bax et al.’s ‘possibly abnormal’ group shows prevalence decreasing from 17 to 12 to 7% between 2 and 4.6 years. This study suggests that there may be a group of children for whom development is especially variable in rate but that may have less entrenched problems. These problems may tend to reduce over time perhaps for no other reason than test/retest error or regression to the mean (Robson 1993). But a further possibility is that this reduction comes as a result of the effect of speech and language therapy services, a point made by Butler (1989) with regard to the Bax et al. study.

Confounding factors

There are potentially confounding factors (most notably, gender, socio-economic status and bilingualism) in the literature that need to be considered. A propensity for marked speech and language delays to be more common in males than females is generally confirmed by the studies reviewed here. Gender ratios derived are 1.25:1 (Randall et al. 1974), 2.26:1 (Stevenson and Richman 1976), 2.30:1 (Burden et al. 1996), 1.25:1 for both speech and language at 4 years (Stewart et al. 1986) and 2.3:1 (speech), with 1.2–1.6:1 (language) (Tuomi and Ivanoff 1977). There are two exceptions to this pattern. One is Beitchman et al. (1986), who found the reverse pattern for speech only (0.98:1), language only (0.98:1) and speech or language (0.82:1), and a most unexpected 0.46:1 for the speech and language diagnosis. The other is Tomblin et al. (1997), who suggest that while boys are more likely to present with SLI, the ratio is nearer equivalence. There are possible explanations for these figures. The first is suggested by the design of the Beitchman et al. study that sought to sample and then project the false-negatives back into the original population sample. Of the false-negatives, the majority was girls and in projecting back up to the main sample the authors projected the gender balance as well as the number of cases. A second explanation is that the relatively liberal cut-off effectively misses the commonly observed discrepancy between the genders because those cases found may be less likely to be true clinical cases and as such may tend to reflect the normal gender balance in the population. A third explanation and one favoured by Tomblin et al. is that existing data are the result of underreporting of difficulties in
girls, a phenomenon which has also been reported in the literature related to reading disabilities (Shaywitz et al. 1990). The fact that the other major cohort study in this area did detect the predicted imbalance (2:1) adopting the fifth percentile as a cut-off suggests that it may be the cut-off which is the determining factor here rather than a high level of undetected difficulties in girls (Silva 1980).

The studies quoted here are not generally helpful in addressing the issue of increased prevalence in lower socio-economic groups. Researchers have commented on this issue but the inclusion criteria for this review effectively excluded groups that might be considered as lower socio-economic status. Some of the studies were carried out in areas with a relatively advantaged population (Burden et al. 1996, Rescorla et al. 1993). Bax et al. (1983) commented on the cumulative effect of low SES on language delay. Similarly Harasty and Reed (1994) give some indication of the potential effect of introducing variation of this type into prevalence estimates. For example, they found higher rates of refusal when making up their diagnostic sample of lower SES children.

Also the data here do not address bilingual or ethnically diverse populations. Stevenson and Richman (1976) deliberately excluded non-indigenous families from their study. Tomblin et al. (1997) noted increased prevalence of language delays in monolingual African-Americans, a finding not replicated by Stewart et al. (1986) in the only reviewed study explicitly to target prevalence in a black population. Wong (1992) found levels of language delay for children from Hong Kong comparable to those in other studies.

In the review no studies were found, with the exception of Tomblin et al. (1997) (also Records and Tomblin 1994), which attempted to integrate clinical judgement and standardized procedure and to determine the effect that this might have on the estimate of prevalence. Accordingly, it is necessary to interpret the above results as being based on a notional psychometric convention. One would argue that this may tend, especially when seen in terms of a cut point of more than 1 SD below the mean, to lead to a relatively liberal cut-off which may, in turn, overestimate real need. However, it does not help one to escape the essentially circular nature of statistically derived prevalence estimates.

Results from natural history studies

A range of different findings is reported, namely those relating to speech and language, language only and speech-only delays. As above, the criteria for determining what is a case is established by the authors of the papers in question, and the disparate nature of some of the findings probably reflects these criteria as much as they do real differences in the groups. This section summarizes each category in turn reporting median figures for each category. Persistence here is defined as the number of children who remain cases of speech and/or language delay at the longest time point quoted by the study. Importantly only speech and language outcomes were used rather than broader educational outcomes or behaviour problems on the grounds that these might result in a more comparable data set.

The median persistence figures for studies examining speech only, language only, and speech and language only are given in table 4. The individual studies are listed with their persistence estimates in table 5. Persistence here refers to the proportion of children whom the authors reported continued to meet their criteria for case status. The median figures for studies examining expressive/receptive,
Table 4. Median persistence reported for speech and language, language only and speech only

<table>
<thead>
<tr>
<th>Type of delay (no. of studies)</th>
<th>Median persistence (%)</th>
<th>Range of reported persistence (%)</th>
<th>Age range (years;months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech only (3)</td>
<td>50</td>
<td>22–54</td>
<td>4;10–33;0</td>
</tr>
<tr>
<td>Language only (10)</td>
<td>66</td>
<td>0–100</td>
<td>0;10–7;0</td>
</tr>
<tr>
<td>Speech and language (1)</td>
<td>38</td>
<td>n/a</td>
<td>3;0–7;0</td>
</tr>
</tbody>
</table>

Table 5. Persistence for individual studies: speech only, speech and language, and language only

<table>
<thead>
<tr>
<th>Persistence Age range Cases/sample size (years;months)</th>
<th>Language (expressive and receptive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech and language Fiedler et al. (1971)†</td>
<td>46/138 38 3;0–7;0 Med</td>
</tr>
<tr>
<td>Median</td>
<td></td>
</tr>
<tr>
<td>Language (expressive and receptive)</td>
<td></td>
</tr>
<tr>
<td>Hall et al. (1993), Hall (1996)</td>
<td>5/9 100 4;7–7;0</td>
</tr>
<tr>
<td>Rescorla and Schwartz (1990)</td>
<td>25/25 54 2;2–3;0</td>
</tr>
<tr>
<td>Richman et al. (1982)</td>
<td>22/705 65 2;0–3;0</td>
</tr>
<tr>
<td>Scarborough and Dobrich (1990)</td>
<td>4/16 0 2;6–5;6</td>
</tr>
<tr>
<td>Silva et al. (1983)†</td>
<td>23/1027 78 3;0–7;0</td>
</tr>
<tr>
<td>Thal and Tobias (1992), Thal and Bates (1998)</td>
<td>10/30 40 1;10–3;0</td>
</tr>
<tr>
<td>Klee et al. (1998)</td>
<td>6/36 67 2;0–3;0</td>
</tr>
<tr>
<td>Ward (1992)</td>
<td>119/321 82 1;0–2;0</td>
</tr>
<tr>
<td>Ward (1992)</td>
<td>61/321 73 0;10–1;10</td>
</tr>
<tr>
<td>Ward (1992)</td>
<td>23/321 50 0;10–1;10</td>
</tr>
<tr>
<td>Median</td>
<td>66</td>
</tr>
<tr>
<td>Speech (6)</td>
<td></td>
</tr>
<tr>
<td>Bralley and Stoudt (1977)†</td>
<td>60/60 22 6;6–11;6</td>
</tr>
<tr>
<td>Felsenfield et al. (1992)</td>
<td>24/52 50 4;10–33;0</td>
</tr>
<tr>
<td>Renfrew and Geary (1973)</td>
<td>150/150 54 5;0–5;6</td>
</tr>
<tr>
<td>Median</td>
<td>50</td>
</tr>
</tbody>
</table>

†Persistence for the longest of three periods within the same study.

Table 6. Median persistence for receptive language only, expressive language only and expressive and receptive language

<table>
<thead>
<tr>
<th>Type of delay (no. of studies)</th>
<th>Median persistence (%)</th>
<th>Range of reported median persistence (%)</th>
<th>Age range (years;months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressive and receptive</td>
<td>75.6</td>
<td>26–100</td>
<td>0;10–7;0</td>
</tr>
<tr>
<td>language (6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive language only (5)</td>
<td>40</td>
<td>0–54</td>
<td>0;10–7;0</td>
</tr>
<tr>
<td>Receptive language only (1)</td>
<td>8.7</td>
<td>n/a</td>
<td>3;0–7;0</td>
</tr>
</tbody>
</table>

expressive and receptive delays are given in table 6, and the corresponding individual studies are listed with persistence estimates in table 7. Median persistence is derived from the single quoted figure for each sample of children with the study concerned.
Table 7. Persistence for individual studies: receptive language only, expressive language only and expressive/receptive language

<table>
<thead>
<tr>
<th>Persistence</th>
<th>Age range (years;months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressive/receptive language</td>
<td></td>
</tr>
<tr>
<td>Richman et al. (1982)</td>
<td>22/705</td>
</tr>
<tr>
<td>Ward (1992)</td>
<td>119/321</td>
</tr>
<tr>
<td>Ward (1992)</td>
<td>61/321</td>
</tr>
<tr>
<td>Hall et al. (1993)</td>
<td>5/9</td>
</tr>
<tr>
<td>Klee et al. (1998)</td>
<td>6/36</td>
</tr>
<tr>
<td>Silva et al. (1983)</td>
<td>23/1027</td>
</tr>
<tr>
<td>Median</td>
<td></td>
</tr>
</tbody>
</table>

| Expressive language | |
| Rescorla and Schwartz (1990) | 25/25 | 54 | 2;2–3;0 |
| Thal and Tobias (1992) | 10/30 | 40 | 1;10–3;0 |
| Scarborough and Dobrich (1990) | 4/16 | 0 | 2;6–5;6 |
| Silva et al. (1983) | 21/1027 | 28.6 | 3;0–7;0 |
| Ward (1992) | 23/321 | 50 | 0;10–1;10 |
| Median | | 40 | |

| Receptive language | |
| Silva et al. (1983) | 23/1027 | 8.7 | 3;0–7;0 |

† Persistence for the longest of three periods within the same study.
‡ Median between two expressive scales reported.

To give some indication of the populations reported in tables 5 and 7, the total number of clinical cases relative to the size of the sample is expressed as case/sample size. The population could have been made up entirely of cases or could have been made up of a larger population sample from which a clinical subgroup was picked out.

**Speech and language delay**

Only one very early study combined speech and language delay (Fiedler et al. 1971) in following children between 3 and 7 years of age. The relatively high figure for spontaneous improvement over the 4 years of the study needs to be contrasted with the much higher level of persistence over the first year (87%). Of interest here is that of the control group 29% developed problems by 4 years and 8% had minor speech problems at 7 years. Unfortunately beyond knowing that the children failed a screening measure at 3 years there is little in this study about the initial presenting symptoms of the children concerned.

**Language delay only**

As with the prevalence data this group represents the largest single source of data. Of the 10 studies, the range is extreme in this group although it is apparent that taking out the studies with the smallest samples would also reduces this range considerably. Thus, if studies with samples below 30 are removed the range shifts to 50–82%, although the median only increases marginally (to 70%). Ward (1992) followed infants from a mean age of 1–2 years of age (tables 5 and 7). The sample
is divided into three groups, being expressive and receptive delay with listening
difficulties, expressive and receptive delay without listening difficulties, and expres-
ptive delay alone. There is a differential outcome for the three groups: while 82% of
the first group continues to show language delay, the figures are 73 and 50%
respectively for groups 2 and 3. Of interest is the fact that the type of presenting
delay changed in some cases from a combined expressive-receptive delay to one of
expressive delay alone.

Speech delay only

Speech problems may appear less persistent than language problems, in that Bralley
and Stoudt (1977) report up to 78% of articulation errors resolving. However, the
long-term data from Felsenfeld et al. (1992) suggest that underlying language dif-
ficulties may continue for children originally identified as having speech delay. Also,
there is a relatively small body of evidence from follow-up studies of children
treated for speech problems (and thus not represented in the present review) to
suggest that literacy skills are at risk even after resolution of speech delay (Stackhouse

The studies looking at speech development give a broad span of 6 months, 5
and 28 years follow-up, all starting with subjects initially ~ 5 years of age. Over 6
months Renfrew and Geary (1973) showed that 54% of speech cases persist, while
Bralley and Stoudt (1977) reported nearly 22% of speech problems persisting over
5 years. Felsenfeld et al. (1992) followed some of a longitudinal cohort (begun in
1960) to track adulthood outcomes. In this study children definitely had no therapy
until third grade (USA schooling, ~ 8 years); there is no account in Felsenfeld
et al.’s paper of any later therapy received. Felsenfeld et al. found 50% of children
with residual speech problems, as assessed by sentence level tests of articulation.
Language measures in adulthood also showed deficits relative to controls, even
though the children were originally identified as speech delayed. More broadly, non-
verbal reasoning and personality scores in adulthood were not significantly different
between the group originally speech delayed and the control group. It should be
noted here that the controls were identified in adulthood and were not matched to
those adults who had originally presented as speech delayed.

Language only—expressive and receptive delay

This group of studies corresponds closely to that reported in table 5 once the
smaller studies focusing on expressive delays have been reassigned, the only differ-
ences being that Hall (1996) remains in and the expressive group from the Ward
(1992) study is taken out. The result is that the median changes little. The Ward
data are made up of two separate groups of children both with expressive/receptive
difficulties, but one with and one without concomitant listening difficulties. In one
study 65% of delays persisted to age 4 (Richman et al. 1982) and 38% of the same
group was experiencing cognitive deficits at 8 years. Of particular interest is Silva
et al.’s data (1983), which show that those children presenting with a language delay
are a fluctuating group; some children failed at each of the three language assessment
points and some failed at only one or two assessment points. There is, however,
the possibility of test-retest error contributing to this effect. The more stable
subgroup comprised those children with generalized language problems affecting both receptive and expressive skills.

*Language only—expressive delay only*

Again there was a considerable disparity between the samples adopted. Two were taken from relatively large-scale population studies (Silva et al. 1983, Ward 1992) and the other three were recruited specifically for the study by notices in paediatricians offices or local newspapers. Some indicated precise criteria for expressive language delay (Rescorla and Schwartz 1990). Otherwise we were looking to identify late talkers on the basis of expressive measures but monitored receptive language and other skills (Scarborough and Dobrich 1990, Thal et al. 1991, Thal and Tobias 1992). The picture that emerges is one of relatively high spontaneous remission. Interestingly, where receptive skills are monitored there is some indication that expressive skills tend to improve but that there is little change in receptive skills (Scarborough and Dobrich 1990). Thal and Tobias (1992) noted a better outcome for those children with normal receptive skills and who use gesture to compensate for their lack of expressive output.

Additional outcomes for areas other than spoken language are given by three studies (Scarborough and Dobrich 1990, Richman et al. 1982, Silva et al. 1987) (table 5 and 7). These all point to reduced reading skills at age 7 or 8 years among those with earlier language delay (whether or not that oral language delay has resolved).

*Language only—receptive delay only*

Only one study sought to identify a group of children with receptive language delay in the absence of expressive delays (Silva et al. 1982). Although it did prove possible to identify such a group the fact that only two of 23 children continued to have problems rather suggests that such children are not at obvious risk for persistent language difficulties. Nevertheless 46% of this group did go on to have reduced IQ and reading performance at age 7. This also suggests that this is a lower risk group than the expressive delay only group if language measures are the target outcome but of comparable risk when other areas of development are taken into consideration. The risk of persistent problems would appear to be much lower than that for the expressive/receptive delay group.

*Summary of major findings*

Inevitably such a review is a retrospective exercise and it is only possible to use the classification framework adopted by the authors concerned. Combining data across studies can be problematic because of potential incompatibility. Nevertheless there are major differences in the range of persistence reported. In the main the figures for speech delay appear to be more variable than those for language delay. Equally there is considerable variability in the persistence of speech plus language delay in untreated samples. The highest level of persistence is that for expressive and receptive language delay, in turn suggesting that it is this group which is recognized and best defined in terms of its case status.
Discussion

At one level there does seem to be a measure of consensus emerging regarding the prevalence of speech and language delay. The median of 5.95% quoted here corresponds well with the 5.59% that was the earliest estimate quoted for speech impairment (Blanton 1916) and with the 6% found by Provonost (NINDS 1969) in a large-scale study (87,288 subjects) identifying speech and language impairments in New England (both cited in Healey et al. 1981). Neither of these works was included in the present synthesis for methodological reasons. Of course reliability does not necessarily imply validity and it is unclear whether the definitions adopted are comparable from a vertical (severity of delay) or a horizontal (range of included measures) perspective. However, this level of convergence suggests that some agreement may be emerging as to what level of difficulty should be considered a clinical problem. In few of the studies reviewed was any attempt to monitor explicitly persistent schooling difficulties. Nevertheless it might reasonably be argued that the level set corresponds to a clinical understanding about which children are most ‘at risk’.

These estimates suggest that speech and language delay in early childhood is a common problem which can have important implications for the child. Whether such delay is, in fact, a sufficiently important health issue to merit population screening is addressed elsewhere (Law et al. 1998). Even a moderate or mild delay may cause appreciable concern to parents and other carers. For a high prevalence condition such as speech and language delay the priority in terms of the screening process is to prevent the identification of false positives (maximize specificity) while maximizing the proportion of true positives (sensitivity). Acceptable levels of accuracy of such measures are discussed in Law et al. (1998) but their validity will, in the end, depend on the relative costs of over and under identification. There are also social and economic costs associated with different outcomes but these have yet to be explored. The natural history data suggest that a predictive element needs to be built into any such programme. To date no such model exists.

The present review indicates that at least two common beliefs are not based on evidence from existing prevalence data. There is little evidence to suggest declining prevalence across the 0–16-year age range, with the possible exception of a sharp drop in expressive delays after 2 years, and there are no data to suggest that the prevalence has increased over the past 30 years. This apparent stability appears to contradict the current understanding of the progression of developmental language delays. In particular, the natural history data suggest that for many children (~60% of children with early expressive delays, perhaps 25% with expressive/receptive delays) difficulties resolve. The explanation for the stability comes from the relationship between prevalence and incidence. If the increase in the number of cases corresponds to the decrease resulting from spontaneous remission the prevalence will remain static. Clearly the key to good prevalence data is longitudinal rather than cross-sectional data. There is no evidence which would suggest that there is a real increase in cases during the 30 years covered by the review (1967–97) despite the fact that many report increased rates of referral and educational placement for these children particularly in the early school years (Jowett and Evans 1996, Reid et al. 1996, SOEID 1996). This suggests that the estimation of prevalence and the demands made on services are not necessarily equivalent. Indeed, referral is likely to be a function of the level of service provision and cannot be equated with
prevalence. Only with a greater understanding of the longitudinal picture of cases can prevalence be established.

The present data set included studies where replicable criteria for case status were provided. In the majority of cases this involved the application of standardized assessment measures. The advantage of this approach is that the resulting data are easier to compare with those from other studies. The disadvantage is that there is an implicit circularity in setting a cut-off on a measure that has been developed on a normal population. Identical cut-offs on standardized measures applied at different ages will necessarily result in the same proportion of the population being identified. Any discrepancy between the proportion of the population predicted from the cut-off adopted and that found in the prevalence study can better be explained by the potential differences between the two populations than by differences in the rate of true cases in the population studied. Cut-offs tend to follow psychometric convention without any direct attempt to link them with clinical judgement. Clinical judgement of case status may also be problematic because it is likely to relate to availability of services and to the anticipated response to therapy but, as Tomblin et al. (1997) found, it is likely to be closely related to psychometric norms and an explicit link between the two may result in higher clinical validity. Stronger support for prevalence estimates would be indicated if the rates derived by clinical judgement of case status showed agreement with those rates derived using conventional cut-off scores.

The review was charged with the identification of gaps in the literature. Apart from the lack of clarity regarding the level of case status there is one major omission in the data. The one important group that has been almost completely ignored in the available prevalence literature related to speech and language disorders is the child with pragmatic impairments. It is true that this group has only relatively recently been identified as communication impaired and it is true that many such children would be subsumed into the autistic continuum. However, given the intransigent nature of many of the problems experienced by these children (Rutter et al. 1992), it is important to know whether they have been incorporated into the other figures or omitted as a distinct non-linguistic category.

The evidence suggests that our understanding of the natural history of speech and language delay, and thus the definition, of case status remains something of an imprecise science at present. Improving prediction should help resolve issues of case definition by demonstrating what would happen if the children were not treated. Good and poor outcome groups could then be traced back to the point at which they were initially identified and relative risk assigned to individual profiles. Furthermore, it should be possible to establish what has been referred to as the ‘critical point’ in the progress of the disability before which therapy is either more effective or easier to apply than afterwards (Sackett et al. 1991). This review confirms the conclusion reached by a number of other studies that children with language delays which incorporate both expressive and receptive skills present the clearest picture. These children are likely to find it difficult to process incoming language, to initiate communication with others and to formulate their responses appropriately. Accordingly they are less likely to compensate for their difficulties and are most likely to find difficulty in coping with the demands of school.

Many of these studies report persistence over comparatively short periods. While this may be interesting in its own right it does not help address the more central question of whether speech and language delays are likely to have a lasting
impact on the child. If the condition is broadly benign it might be anticipated that persistence would decline the longer the follow-up period. Alternatively it may be that the condition left untreated leads to a wide range of social and educational sequelae. These issues were not fully explored in the present synthesis but are highly relevant to the area of speech and language delay. Whether these oral language delays have resolved, multiple educational and social difficulties are noted for children who had earlier speech/language delays. Of early expressive language delayed children, 41–75% showed reading problems at age 8 years. This finding is confirmed by the larger body of literature from follow-up studies of children who may have received intervention for speech and language delay and who were subsequently identified as being dyslexic (e.g. Catts 1991).

Indeed there is also the issue of the prognosis for children with speech and language delays who have received intervention. Follow-up studies tend to paint a more negative picture of outcomes for children whose difficulties do not resolve by school entry (Petrie 1975, Bishop and Edmundson 1987, Cook et al. 1988, Huntley et al. 1988, Urwin et al. 1988, Haynes and Naidoo 1991, Davison and Howlin 1997). The findings reveal that in addition to continuing problems in verbal language, reading, spelling and other aspects of educational attainment can also be affected (Aram and Nation 1980, Stark et al. 1984, Bishop and Adams 1990, Catts 1991, Morris-Friehe and Sanger 1994, Tallal et al. 1997) together with behaviour and other aspects of psycho-social adjustment (Baker and Cantwell 1987, Silva et al. 1987, Beitchman et al. 1989, 1996). While the problems are more marked for those children whose language difficulties are associated with low intellectual ability or those with both receptive and productive language difficulties, those with primary delay can also experience marked long-term difficulties which may persist into adulthood (Haynes and Naidoo 1991). How does this rather more negative picture square with the natural history data reported above? The main reason for this difference is that these follow-up studies in the main report data on children who are already confirmed ‘clinical’ cases and are, therefore, already in the highest risk group. By including only the most severe cases they may overestimate the negative sequelae of the population of speech and language delayed children as a whole. Given the level of persistence in the treated group it might be more appropriate to argue that spontaneous improvement in this group is no longer the central issue. Rather the outcome should be how well the children can cope with the disabilities associated with their impairments.

The nature of the concept of natural history needs to be addressed. While it may be technically possible to provide no input at all, in reality the needs of the children contribute to the world around them. They receive help from parents, from teachers and nursery staff, which may reflect those needs. Intervention of this type may not be comparable to specialist speech and language input but it is difficult to say that it has no effect. The term ‘natural history’ is derived from less context-dependent interventions. Thus, a drug regimen may be said to be in place or not. If it is not, it may be possible to follow the course of the disease. However, it is not clear that the same is true of speech and language delays especially in children who, as a group, may be paid particular attention by all those around them by virtue of those very difficulties. A second difficulty in interpretation is the ethical implications of withholding intervention in order to carry out such studies. While this may have been true when no support services were available, for example for the children in the Dunedin study (Silva et al. 1987), this is now not the case for
most children in the English-speaking world at least. On what grounds would it be reasonable to withhold intervention for the sort of length of time such that it would be meaningful to ascertain natural history? It has been suggested that 3 years is an appropriate period over which to measure change (Aylward 1988). To a certain extent the answer to this question depends on what is known about intervention. On the one hand if one cannot show that intervention is effective it may be that natural history is a reasonable concept. On the other hand if benefits can be demonstrated then the case for withholding treatment over the long-term recedes. The intervention data reviewed in the present synthesis (Boyle et al. 1999) has suggested that there is a good case for assuming that intervention can be shown to be beneficial although the data set remains comparatively small.

This has led some commentators to suggest that a distinction be drawn between routine intervention and specialist intervention (Shriberg 1994, Shriberg et al. 1994). Community provision of speech and language therapy could constitute natural history in the sense that children receive other services within the healthcare system, for example the services of a family doctor or a community nurse. Such services are often of a lower level of intensity than specialist provision and might reasonably be construed as the normal activities of social or health services. However, such an approach does not really help overcome the essentially rather arbitrary distinctions that are drawn between different levels of input. Where does specialist intervention begin—once children have been referred to specialist provision? The move towards integrating children into their local community schools may mean that fewer children would be receiving the kind of specific intervention that would preclude them from involvement in natural history studies of the kind reviewed here.

A model of predictive risk

Interpretation of the data is dogged by an uncertainty as to which children are truly cases in the sense that without intervention they would necessarily go on to experience problems associated with early speech and language delay in school and beyond. This issue is complicated further by social constructs of ‘normality’ and the difficulty in establishing a true ‘gold standard assessment’ with a theoretically determined definition of case status. An alternative approach is to move away from the model adopted in the above studies and towards a model of predicted risk. This would involve identifying the key predictive factors of subsequent case status and to build them into a model that identifies an ‘at risk’ group. This is a relatively new concept in the field of speech and language therapy, although see Tomblin et al. (1997) for some early work in quantifying that risk with relation to SLI. Much depends on the outcome measures adopted. It seems likely that speech and language measures alone will not provide a meaningful outcome and literacy and other educational measures will need to be built into such a model (Beitchman et al. 1996, Drillien et al. 1988, Glascoe 1999). Such an approach would circumvent the need to carry out separate prevalence studies in bilingual or disadvantaged populations but it would mean gathering data, which included representative populations, rather than attempting to model outcomes on a discrete clinical population as is often the case. It would not be a matter of reporting, as most authors have already done, that males are over-represented in this group of children but of establishing to what extent gender is a risk factor in different subcategories of the population. The advantage of such an approach is that it could make use of the available genetic
Prevalence and natural history of primary speech and language delay

(Bishop et al. 1995), and intervention data (Boyle et al. 1999a, b) along with data regarding co-morbidity such as behaviour (Stevenson 1996) as it becomes available to modify and improve the model. Similarly such a model could begin to test the findings of contributory factors identified in other studies. Effectively a model based on risk would be one that embraced the variance in the population.

Early attempts to model the progression of speech and language disorder have relied heavily on regression analysis (Schery 1985) and have often included essentially normal middle class children in their samples (Bee et al. 1982). However, there is a need to look at the capacity of independent variables to discriminate clinical outcomes. To date there have been a great many studies examining the association between other factors and language delay/disorder (Lassman et al. 1980). But with three exceptions little attention had been paid to the risk attached to these associations (Drillien et al. 1988, Tomblin et al. 1997, Paul and Fountain 1999). Tomblin et al., in a retrospective study, adopted the odds ratio metric, a means of ascertaining the level of association between the occurrence of a particular condition and a particular risk factor. The odds of having an exposure to a given risk among clinical cases is contrasted with the odds of having it in the normal population. They calculated the odds ratios associated with case status of SLI at 5 years and found that paternal rather than maternal history of speech and language difficulties and paternal smoking during foetal development, were particularly associated with specific language difficulties. Conversely breast-feeding and length of breast-feeding functioned as a protective factor effectively reducing the risk of SLI. Paul and Fountain (1999) examined a small prospective cohort of thirty six children with specific expressive language delays identified at 20–34 months and followed up to 7 years of age. Using discriminant function analysis, they found that the best predictors of outcome were socio-economic status, parent report of adaptive behaviour and gross motor skills. Although case status in these studies was comparable (10th centile) the measures differed as did the variables put into the analyses which were in themselves based upon different statistics. In neither case was the level of intervention factored into the analysis.

It is also possible to look at speech and language delays as markers for other conditions (Whitehurst et al. 1991). Drillien et al. (1988), for example, used early language performance as predictors of subsequent language and educational development. They examined the relationship between developmental screening results in five developmental areas and their association with subsequent educational and language performance. They reported significant odds ratios of 1.60 (CI 1.12–2.27) to 2.44 (1.5–3.97) for the association between all aspects of developmental screening and subsequent expressive language difficulties but found much weaker associations between early expressive language difficulties and subsequent school performance. This corresponds well with the relatively low level of persistence of expressive difficulties reported in the present review. Each of these associations was calculated on the basis of the individual child’s performance at developmental screening checks between 20 weeks and 5 years. Like Paul and Fountain they emphasized the saliency of adaptive behaviour and motor skills in the predictive model.

The aim of such analyses would be to provide the clinician with the tools to say that, given the following presenting symptoms, and predisposing and ameliorating factors the following outcome could be predicted given the following intervention. For example, it might be possible to say that a child with an expressive delay has an even chance of developing persistent effects either in language or in
other skills. These odds may lengthen if negative social interaction and medical factors are included but may shorten if short-term intervention of known impact is added into the equation. For example, Boyle et al. (1999a, b) indicate a mean effect size of the order of 1 SD for relative short-term ‘dosage’ interventions for expressive language delays in the pre-school years. Conversely, it might be possible to say that a child with an expressive/receptive difficulty may have a seventy five percent chance of developing associated difficulties, for example a risk which may be increased if other contributory factors are added but which may not be alleviated with short-term intervention. Finally there may be differential effects of impairments and environmental effects across time. For example, there is some suggestion that for children with motor difficulties the motor skills serve as good predictors in the short-term but the environment plays as a more substantial role as the child gets older (Cohen et al. 1986). The role of social class is clearly one that needs to be examined with care. Across a population it is likely to operate as a proxy variable accounting for a considerable proportion of the variance. However, in doing so it is likely to mask the effects of other components in any analysis, components, which may be more amenable to modification at the level of the individual.

Reviews of the literature often end with a call for more research. The conclusion drawn here is that while health service providers may wish for measures of prevalence upon which to base services it would be premature to call for more prevalence data. More population estimates may be warranted particularly in relation to the needs of local populations but it is the risk not the single prevalence estimate towards which progress most needs to be made. There is a need for large-scale cohort studies to tease out the relationship between the component parts of the equation across time and to establish a better understanding of the incidence of speech and language delays. More attention needs to be paid to the risk and protective factors and the extent to which they are influenced by confounding variables and their interactions.

Conclusions

This paper has demonstrated the close association between prevalence and natural history. Appreciation of one cannot follow without a good understanding of the other. Ultimately the issue revolves around the concepts of prediction and relative risk. Such issues can only be properly dealt with using longitudinal population data and with clinical validation of psychometric classification.

Acknowledgements

The NHS Centre commissioned this systematic review for Reviews and Dissemination at the University of York. F. H. received support from the Department of Clinical Communication Studies, City University, London, in the latter part of this project. The authors also thank Christopher Norris, Biomedical Research Indexing, and two anonymous reviewers of the paper.

References

Prevalence and natural history of primary speech and language delay


Sheridan, M. D. and Peckham, C., 1975, Follow-up at 11 years of children who had marked speech defects at 7 years. Child: Care, Health and Development, 1, 157–166.


Prevalence and natural history of primary speech and language delay


Appendix 1: Studies meeting the criteria for inclusion in the prevalence domain of the review


Thorburn, M. J., Desai, P. and Durkin, M., 1991, A comparison of efficacy of key informant and


**Appendix 2: Studies meeting the criteria for inclusion in the natural history domain of the review**


**Thal, D. J. and Bates, E., 1988, Language and gesture in late talkers.** *Journal of Speech and Hearing Research*, 31, 115–123.

