Chapter 3. Results

Key question 1: What is the Natural History (Spontaneous Resolution Rate over Time without Treatment) for

- OME persisting after a discrete episode of acute otitis media,
- newly diagnosed OME of unknown duration (unilateral or bilateral),
- OME persisting for weeks or months (unilateral or bilateral),
- unilateral OME lasting 3 months or longer,
- bilateral OME lasting 3 months or longer?

Literature Review

After initial screening of 4,879 titles or abstracts, we identified 449 articles for review. After secondary screening of the 449 articles, we identified 141 articles that fell within the scope of this question. Tertiary screening identified 38 articles on prospective cohort studies for potential abstraction. After reassessing articles included in the systematic review by Rosenfeld (1999), we included three more articles for potential abstraction. We eliminated one article on r OME following acute otitis media which was addressed in a recent evidence analysis. A total of 26 prospective cohort studies and one retrospective-prospective cohort study were identified among these 40 articles. **Table 17** lists the studies and cohorts examined. Abstraction was possible from 33 of the 40 articles. **Table 18** lists the studies referred to by Rosenfeld (1999) and their disposition in the present evidence-based analysis. **Table 19** lists the articles excluded because relevant data could not be abstracted.

Findings

Evidence Table 1 presents the study characteristics, population characteristics, risk factors, and findings for the 27 cohort studies described in 34 articles that responded to this question**.** **Table 20** presents the study quality scores for the studies included in **Evidence Table 1**. The quality scores (see Methods) for these studies fell at the low end of the possible range of 1(lowest) –to 6(highest). Three of the cohort studies had a score of 4, sixteen had a score of 3, seven had a score of 2, and one had a score of 1.

Although we accepted these 27 studies for the natural history analysis, half of them (thirteen) failed to document that the subjects had received no medical or surgical treatment during the course of the study that could affect the outcome of OME. In the three studies that claimed that their subjects received no treatment, the investigators failed to document how adherence was maintained or confirmed. In studies of eleven cohorts, the authors mentioned that children received antibiotics or underwent surgical procedures that could affect OME outcome. In those studies that reported numbers of children who received treatment, the proportion was relatively

small compared to the total number studied. Except in two studies, results were not stratified by treatment condition, even in those articles that reported treatments received by study subjects. Further, the majority of studies, but not all, used tympanometry as the sole diagnostic test of OME. However, the criteria for each tympanogram type in these studies are similar (**Table 21**). Table 21 lists the pressure and immittance parameters for tympanogram types for those studies that used tympanometry as the sole diagnostic criteria for OME.

Tables 22 and 23 provide OME resolution rates by ear, as reported in the majority of the studies, and by child for those studies in which data could be stratified by unit of analysis (ear vs. child), age group, and OME type respectively. For clarity, we have listed only those resolution intervals with onset at the study's inception. For clarity, when tympanometry was the diagnostic method, counts for tympanogram type B transition to A are shown when provided, while other tympanogram transitions are shown in **Evidence Table 1.**

Data from a number of studies were reported in a format that made them unusable for quantitative syntheses. Ten studies were eliminated from quantitative syntheses, because they failed to stratify by age (i.e. less than and greater than 3 years old). Day-to-day variability in tympanogram types was described in a cohort of kindergarten children examined on each weekday for 30 days (Moller and Tos,1990). Ernston and Sundberg (1984) described a group of children who participated in a controlled trial of children with OME that persisted for at least 3 months; they found that 15.3 percent (11 of 72) of such children had OME resolution by five weeks followup. A study of children with OME that persisted for 3 months or more, showed 45 percent (49 of 109) OME resolution at 2.5 year followup (Leiberman and Bartal,1986). However, the investigators acknowledged that middle-ear effusion noted after such a long interval could be either persistent or recurrent. Birch and Elbrønd (1984) and Zielhuis, Rach, and van den Broek (1990) derived equations to describe the OME resolution rates they observed in their cohorts, but we were unable to abstract actual counts from these articles.

We performed two sets of meta-analyses that matched age groups, unit of analysis, outcome type, and time to resolution in three or more cohorts. All meta-analyses presented here used the 'ear' as the unit of analysis. Few studies considered the child or the episode as the unit of analysis and no meta-analyses were possible.

Resolution at 6-Week Followup The first set of meta-analyis contains two meta-analyses (**Table 24**). The meta-analyses showed that if the criteria for resolution were tympanogram type B or C transition to A or by otoscopy, 42.3 percent (95% CI: 24.1%, 60.6%) of ears with newly diagnosed OME of unknown duration in children older than 3 years had resolution by the 6-week exam. If the criteria for resolution were modified to tympanogram type B transition to A or otoscopy, the proportion of ears in children older than 3 years old with resolution at 6-week followup was 37.2 percent (95% CI: 1.8%, 72.5%). Spontaneous resolution rates were significantly different among the cohorts for both definitions. In these studies, the OME resolution rates were calculated by determining the proportion of children without OME at followup, whether or not their baseline OME had resolved and recurred at an earlier point. Thus, these are not cumulative resolution rates (Sly, Zambie, Fernandes et al., 1980; Lamothe, Boudreault, Blanchette et al., 1981).

Resolution at 3-Month Followup The second set of meta-analyses are shown in **Table 25**. The first two meta-analyses assessed resolution of OME by ear by the time of 3-month followup in children older than 3 years. If the criteria for resolution were tympanogram type B or C transition to A, the proportion of ears with resolution of newly diagnosed OME of unknown duration was 42.7 percent (95% CI: 29.3%, 56.1%) among children older than 3 years. If the

criteria for OME resolution were tympanogram type B transition to A, then the proportion of ears with resolution was 22.5 percent (95% CI: 5.9%, 39.0%). Spontaneous resolution rates were not significantly different among the studies in either comparison, except for Tos, Holm-Jensen, Sörenson, and Mogensen (1982) which had lower resolution rates. Because Fiellau-Nikolajsen and Lous (1979), Fiellau-Nikolajsen (1979), and Renvall, Anniansson, and Lidèn (1982) terminated followup whenever a child had a type A or normal tympanogram and because Tos, Holm-Jensen, Sörenson, and Mogensen (1982) provided data for calculation of the cumulative OME resolution rate, this estimate represents a cumulative resolution rate. The last three meta-analyses in Table 25 show the derivation of the cumulative resolution rates.

Resolution Rates for Younger Age Groups Only two studies examined the resolution rates for each of the age groups of less than 6 months and 3-months to 3-years (see **Tables 22 and 23**). Thus, no meta-analyses for children under 3 years of age were performed.

The Role of Influencing Factors in Resolution Similarly, because very few studies assessed the role of factors that might influence resolution, no meta-analyses were performed. The results of individual studies are summarized here:

- Lamothe, Boudreault, Blanchette, and colleagues (1981) and Fiellau-Nikolajsen and Lous (1979) assessed the effects of gender and found quicker resolution among females than males.

- Fiellau-Nikolajsen (1979) noted that children who received at-home care had quicker resolution of OME than did children in daycare.

- Sly, Zambie, Fernandes, and colleagues (1980) compared OME resolution in small cohorts first studied in February with those first studied in September and found resolution to be more rapid in the February cohort.

- Lamothe, Boudreault, Blanchette, and colleagues (1981) also assessed the effect of the side of the affected ear and noted more rapid resolution in affected right ears than affected left ears.

- Portoain-Shuhaiber and Cullinan (1984) stratified by racial/ethnic origin and noted quicker resolution of OME in African children than in Indian or Caucasian children.

- Moller and Tos (1990) found that different tympanometry instruments, which they described as impedance tympanoscopy and impedance audiometry, gave different rates of OME among the same group of children.

- Zielhuis, Rach, and van den Broek (1990) found that season and age at the end of the episode had a statistically significant effect on OME resolution rates, while gender, upper respiratory tract infection, and history of AOM did not, although they presented only percentages without denominators.

Summary

We found sufficient data only to analyze the rates of resolution of OME among children older than 3 years old. No data existed on the role of such factors as duration of OME, initial occurrence of AOM, or presence of OME in one vs. both ears. Data on the roles of gender, at-home vs. day care, race/ethnicity, time of year, age of onset, or side of ear were too limited to draw any conclusions.

Key Question 2: What Are the Effects of Early-Life OM on Long-Term Speech and Language Development?

- Do infants and preschool children with longer-duration early-life OME as compared to those with shorter duration OME have greater delays in speech and language development (receptive or expressive) later in life?

- Is OME-associated conductive hearing loss in the first 3 years of life a risk factor for speech and language developmental delays?

- What are the risk factors that interact with the effect of OME on speech and language development in infants and preschool children?

Literature Review

After secondary screening of the 449 articles retrieved for review, we identified 112 articles that fell within the scope of this question. Tertiary screening identified 20 studies that fulfilled the 5 criteria for analysis. (Table 26) The five criteria included the following:
1) OM was diagnosed/assessed before the age of 3 years, 2) speech or language outcome was measured at or before the age of 22 years, 3) a prospective cohort study design was used, 4) OM was graded, and 5) speech or language outcome was measured beyond 3 years of age. Of the 20 studies, 17 were prospective cohort studies and three (Freeark, 1992; Fischler, 1985; Paul, 1993) were retrospective-prospective studies in which OM history was retrospectively obtained, but outcome measures were prospectively obtained. We did not exclude any studies based on their quality.

When we reviewed the studies included in the 1994 OME Guideline for inclusion in our analysis (Stool, Berg, Berman et al., 1994), we excluded six of the studies. The Friel-Patti and Finitzo (1990) study was excluded, because it was not a prospective cohort study and did not report outcomes for children over 3 years of age. The Friel-Patti (1982) study was excluded because no outcomes were measured in children over 3 years of age. The studies of Lous and Fiellau-Nikolajsen (1988) and Rach, Ziehlhuis, van Baarle, and colleagues (1991 were excluded, because the investigators did not measure OM severity before the age of 3 years. The Rach, Zielhuis, and van den Broek (1988) study was excluded, because it was not a prospective cohort study. The study by Wright, Sell, McConnell, and colleagues (1988) was excluded, because no outcomes were measured in children over 3 years of age. Finally, unlike the OME Guideline (Stool, Berg, Berman et al., 1994), we included the studies by Klein (1988) and Freeark (1992), because study quality was not one of our rejection criteria.

Table 27 lists the author, year and cohort of 22 prospective cohort studies that were excluded from our analysis because they did not report outcomes for children older than 3 years of age.

These 22 studies actually included only 12 cohorts. Relevant findings for seven of these 12 cohorts were included in our assessment.

Findings

Evidence Table 2 presents the study characteristics, population characteristics, risk factors, and outcome findings for the 20 cohort studies that responded to this question. Speech and language outcomes were examined in a total of 12 cohorts of children. Of these 12 cohorts, nine included children primarily from specific populations, such as a particular ethnic or racial group or a particular socioeconomic group. It is also important to reiterate that cohorts were excluded if they consisted exclusively of children with craniofacial defects, primary mucosal disorders, immunodeficiencies, or a genetic disorder. None of the studies specifically assessed children who already had speech, language, or other developmental delays. One of the studies focused specifically on persistent bilateral OM.

For the purpose of responding to this question, we defined early life otitis media as positive otitis media history prior to 3 years of age. **Table 28** summarizes the definitions of positive or negative history of otitis media used in these studies and the diagnostic method for OM. The definition of positive or negative OM history varied among the studies. Some studies used percentage of visits during a specified time, some used number of visits with OM, some used days spent with effusion, while others combined various criteria.

The age during which the outcome was measured also varied. This age ranged from one to 3 years. The diagnostic method and the examiner also varied. Several studies based diagnosis on chart review or parents' record. A few studies used pneumatic otoscopy performed by pediatricians, otolaryngologists, or trained professionals.

The quality of the studies included in the evidence table is summarized in **Table 29**. Of the 20 studies, five (25 percent) received a score of six of a possible of eight points; four (20 percent) scored five points; six (30 percent) scored four points; four (20 percent) scored three points, and one (5 percent) scored two points.

Table 30 presents a summary of the key characteristics of the 20 cohort studies including the cohort, age of OM history, age at outcome measure, major outcome statistical analysis, outcome measure of interest, diagnostic procedure(s), and notes. The table is organized by the outcome measures for this key question, namely cognition, expressive language, receptive language, expressive speech, and receptive speech. Since multiple outcomes could be included in one study, a study may appear in several rows. The entries in the "Test" column should be interpreted with caution, because subtests versus global tests were not distinguished.

The factor we used to determine which studies to examine further by meta-analysis was the major outcome measure, which indicates the type of statistical measured used. Whenever three or more studies reported the same outcome measure, we considered pooling the data. Based on this strategy, we conducted three possible meta-analyses to derive the pooled difference between positive and negative early-life otitis media for: expressive language development, receptive language development, and development of cognitive verbal intelligence.

Expressive Language Development **Table 31** presents the findings of the three cohort studies that addressed expressive language development. All three studies measured OM history prior to 3 years of age. Although the pooled standardized mean difference showed an increase of about 14.5 percent (95% C.I.: -49.2%, 20.2%) of a standard deviation in the expressive language

measure for the group of children with no early-life OM history compared with those with a positive history, this pooled estimate is not significantly different from zero. Thus, the available data do not support the hypothesis that an OM history prior to 3 years of age has an effect on expressive language development. However, the 95% confidence intervals on our pooled results do not exclude a clinically important effect size of almost 0.5, meaning no strong conclusions can be drawn. The Chi-squared test of heterogeneity showed that the standardized difference was not significantly different among the studies. However, the ages at which outcome was measured and the tests used were not uniform across the studies.

Receptive Language Development **Table 32** presents the findings of the four cohort studies that addressed receptive language development. All four studies measured OM history at less than 3 years of age. Although the pooled standardized mean difference showed an increase of about 10.3 percent (95% C.I.: -28.9%, 49.5%) of a standard deviation of the receptive language measure in the group of children with no early-life otitis media history, this pooled estimate is not significantly different from zero. Thus, the available data do not support the hypothesis that an OM history prior to 3 years of age has an effect on receptive language development. However, the 95% confidence intervals on our pooled results do not exclude a clinically important effect size of almost 0.5, meaning no strong conclusions can be drawn. The Chi-squared test of heterogeneity showed that the standardized mean difference was not significantly different among the studies. However, the age at which outcome was measured and the test used were not uniform across the four studies. In addition, although the racial/ethnic composition was not reported precisely, one of the studies included primarily African-American children, another primarily American Indian children, and the third primarily Caucasian children in private practice.

Cognitive Verbal Intelligence **Table 33** presents the findings of the three cohort studies that addressed development of cognitive verbal intelligence. All three studies examined cognitive verbal intelligence, and all measured OM history at less than 3 years of age. Although the pooled standardized mean difference showed an increase of about 23 percent (95% C.I.: -20%, 65%) of a standard deviation in the expressive language measure in the group of children with no early life OM history, this pooled estimate is not significantly different from zero. Thus, the available data do not support the hypothesis that an OM history prior to 3 years of age has an effect on the development of cognitive verbal intelligence. However, the 95% confidence intervals on our pooled results do not exclude a clinically important effect size of almost 0.5, meaning no strong conclusions can be drawn. The Chi-squared test of heterogeneity showed that the standardized mean difference was not significantly different among studies. However, the age at which outcome was measured and the test used were not uniform across the studies. Further, two of the study populations were primarily African-American. The third study population was of lower socioeconomic status.

Summary

The data do not support an effect of early-life OME on language development or cognitive verbal intelligence. However, differences among the cohorts and study conditions and the wide

95% confidence intervals make it difficult to conclude that there is no effect. We found insufficient data to assess early-life OME on speech development.

Key Question 3: What are the Effects of Early-Life OM on Long-Term Hearing?

- Do infants and preschool children with longer duration early life OME as compared to those with shorter duration OME have permanent (or sensorineural) hearing loss later in life? One specific formulation of this question is: Is OME-associated conductive hearing loss in the first 3 years of life a risk factor for permanent (or sensorineural) hearing loss later in life?

- What are the risk factors that interact with the effect of OME on hearing later in life (unilateral or bilateral) in infants and preschool children?

Literature Review

After secondary screening of the 449 articles we retrieved for review, we identified 186 articles that fell within the scope of this question. Tertiary screening identified 12 studies that fulfilled the five criteria for analysis. The five criteria included 1) OM was diagnosed/assessed before the age of 3 years, 2) hearing outcome was measured at or before the age of 22 years, 3) a prospective cohort study design was used, 4) OM was graded, and 5) hearing outcome was measured after 3 years of age. Of the 12 cohort studies, four were excluded from further analysis. The reasons for exclusion of the four studies are presented in **Table 34**. **Table 35** lists the eight studies included in the evidence table and considered for analysis. **Table 36** lists the author, year, and cohort of 10 studies that were not included in the analysis because they failed to report findings for children over 3 years of age.

Findings

Evidence Table 3 presents the study characteristics, population characteristics, risk factors, and findings of the eight cohort studies considered for this question.

Table 37 summarizes the definitions of positive and negative history of OM used in these studies and the diagnostic method for OM. The definitions of positive and negative OM history varied from one study to another. These variations were similar to those identified in the studies of speech and language development.

The study quality of the eight studies included in **Evidence Table 3** is summarized in **Table 38**. Of the eight studies, 1 (12.5 percent) received a score of 6 of a possible of 8 points; 4 (50 percent) scored 5 points; 2 (25 percent) scored 4 points; and one (12 percent) scored 2 points.

The age at which OM history was taken, age at outcome measure, and type of outcome measure for the 8 cohort studies are displayed in **Table 39**. With the various combinations of age at outcome and type of outcome measure that characterized the studies, only one combination, percentage of conductive hearing loss at 6 to10 years of age, was considered sufficiently clinically similar to justify statistical pooling: four studies reported this outcome

measure at this age range. Three of the four studies (Fischler 1985, Harsten 1993 and Sorri 1995) reported treatment, including oral antibiotics, myringotomy, and tympanostomy tube, for OM episodes; Kaplan (1973) did not address treatment. Sorri (1995) used 20 dB as the air-conduction threshold above which hearing loss was defined. The other three studies used 25 dB as the threshold.

We conducted two meta-analyses, one including all four studies and another that excluded the Sorri 1995 studies, which used a different threshold for hearing loss from that of the other three studies. The meta-analysis findings are reported in **Table 40**. The pooled risk of conductive hearing loss among 346 children who had early-life OM was 22 percent (95% CI: 7% to 36%), compared with 6 percent (95% CI: 1% to 12%) among 237 children who did not have a history of early-life OM. The pooled difference in rate of hearing loss between those with an early-life OM history and those without was 11 percent (95% CI: 3% to 19%), and the pooled risk ratio was 2.6 (95% CI: 1.6 to 4.2). Thus, an early-life history of OM was significantly associated with conductive hearing loss. Neither the studies pooled for the rate difference nor the studies pooled for the risk ratio were statistically heterogeneous. **Figure 1** presents the shrinkage plot and **Figure 2** presents the funnel plot for the rate difference and **Figure 3** and **Figure 4** present similar plots for the risk ratio of hearing loss. These figures show that negative early-life OM history is more favorable in term of conductive hearing loss at age 6 to 10 years. The funnel plots for the risk difference and the risk ratio (**Figure 2** and **Figure 4,** respectively) showed no indication of publication bias. Neither the adjusted rank correlation tests (Begg, 1999) nor the regression asymmetry test (Egger, 1997) indicated publication bias for either statistic ($p>0.99$ for the risk difference; $p=0.31$ for the risk ratio; $p=0.71$ for the risk difference; $p=0.28$ for the risk ratio; respectively). However, these results should be interpreted with caution. Only four studies were included in the analysis. Moreover, each of these cohorts included relatively homogeneous populations of children, one from Finland, another from Sweden, one primarily of American Indian children, and another primarily of Eskimo children. The sensitivity analysis that excluded the Sorri (1995) study did not change the conclusions. However, the exclusion of the Sorri study reduced the variability of the rate difference and lowered the pooled difference in rate of conductive hearing loss between those with a positive OM history and those without from 11 percent (95% CI: 3% to 19%) to 8 percent (95% CI: 4% to 13%).

Summary

The results support that history of early-life OME is associated with increased risk for conductive hearing loss. However, the number of studies with similar outcome measures is small. We found insufficient data to assess the early-life OM on permanent hearing loss.

Key Question 4: Diagnostic Methods for OME

What are the sensitivity, specificity, and predictive values for alternative methods of diagnosing OME compared with one of the reference standards?

These methods include, but are not limited to:

- signs/symptoms
- non-pneumatic otoscopy
- pneumatic otoscopy, validated or non-validated examiner
- binaural micro-tympanoscopy
- portable tympanometry
- professional tympanometry
- quantitative tympanometry
- acoustic reflectometry (specify model and year)
- otoacoustic emissions
- audiometry, air or bone conduction thresholds.

Literature Review

After secondary and tertiary screening of the 449 articles we retrieved for review, we identified 75 articles that fell within the scope of this question. When we compared our list with the 1994 OME Guideline (Stool, Berg, Berman et al., 1994), we found five studies that were included in the 1994 OME Guideline but not in our assessment. We excluded the Kaleida (1992) and the Shurin, Pelton and Finkelstein (1977) studies, because data were not abstractable. The McDermott and Giebink, Le, and colleagues (1983) and the Teele and Teele (1984) studies were excluded, because they did not address the scope of this question. The Lampe, Weir, Spier, and colleagues (1985) study was excluded, because it was a duplicate of another study. We included three studies that were rejected by the developers of the 1994 OME Guideline, because we did not reject any studies based on study quality. These were the studies by Haughton (1977), Karma (1989), and Marchart (1986).

Of the 75 articles accepted for data abstraction, we included 52 studies in our assessment (listed in **Table 41)**. **Table 42** lists the reasons for exclusion of the 23 remaining articles. **Evidence Table 4** presents the study characteristics, the characteristics of the study population, and the study findings of each of the 52 studies included.

Of the 52 studies, 33 (63 percent) scored three points or fewer on our six-point quality scale. Of the 19 studies that scored more than three points, 15 studies scored four points, three studies scored five points, and one study scored six points. Most of the studies in this group are of poor quality, a finding that replicates those of Lijmer, Mol, Heisterkamp and colleagues (Lijmer, Mol, Heisterkamp et al., 1999).

Findings

Evidence Table 4 presents the sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and prevalence of OME in the cohort for each comparison of diagnostic methods and reference standards listed within the scope of this assessment.

Table 43 summarizes the number of comparisons for each diagnostic method and reference standard pair. On the basis of these numbers we selected groups of three or more studies for meta-analysis, from which we derived pooled random effect estimate, 95% confidence intervals, and measure of heterogeneity for sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and prevalence.

Tables 44 through **52** present the results of the meta-analyses that compare sensitivity, specificity, and prevalence rate for acoustic reflectometry at <=5 or >5 RU (reflective units), pneumatic otoscopy, portable tympanometry, professional tympanometry using acoustic reflex at 500 or 1000 Hz, professional tympanometry using static compensated acoustic admittance at 0.1, professional tympanometry using static compensated acoustic admittance at 0.2, professional tympanometry using static compensated acoustic admittance at 0.3, professional tympanometry using B curve as abnormal, and professional tympanometry using B or C2 curves as abnormal, respectively, using myringotomy as the reference standard. Findings excluding duplicates are summarized in **Table 53**. The receiver-operator characteristic points that correspond to sensitivity versus (1-specificity) are plotted in **Figure 5.** The receiver operator characteristic points showed that pneumatic otoscopy was closest to the optimal operating point where both sensitivity and specificity would be 100 percent.

Among the nine diagnostic methods, pneumatic otoscopy and professional tympanometry (using flat or B or C2 curve as abnormal) had the highest sensitivity at 93.8 percent (95% CI: 91.4, 96.3) and 93.8 percent (95% CI: 91.1, 96.4) compared with myringotomy, respectively. The diagnostic test with the highest specificity was professional tympanometry (using static compensated acoustic admittance at 0.1) at 95.0 percent (95% CI: 88.5, 100).

If we consider both sensitivity and (1 minus specificity) in the receiver-operator characteristic display in **Figure 5**, pneumatic otoscopy is closest to the optimal operating point where both sensitivity and specificity are 100 percent. The pooled sensitivity was 94 percent (95% CI: 91%, 96%), and the pooled specificity was 80 percent (95% CI: 75%, 86%). These findings were based on 2,694 children from 7 studies that reported a pooled prevalence of OME of 63 percent (95% CI: 58%, 67%). The estimated prevalence rates ranged from 56 percent to 71 percent, which indicated significant heterogeneity ($p<0.001$). We used the pooled sensitivity and specificity for pneumatic otoscopy and derived the positive and negative predictive values for various prevalence levels. **Figure 6** provides such a plot.

Table 54 shows an analysis of the study quality of the diagnostic tests included in the meta-analyses. With the exception of the study by Babonis (Babonis, Weir, Kelly, 1991), which scored five, all studies scored four or less out of a maximum of six possible points. The majority of the studies did not fulfill criterion four on representativeness of patient sample in clinical practice or criterion five on determination of reproducibility of test results. Among the seven studies used to derive the pooled estimates of sensitivity and specificity for pneumatic otoscopy, two studies scored the minimum of one point, two scored two points, one scored three points, and two scored four points.

In **Table 54,** we also compare the qualifications of the examiner performing the diagnostic test for each study. Of the seven studies in the comparison between pneumatic otoscopy and myringotomy, two studies did not specify the test performer, one study specified that a senior registrar and a senior house officer performed the test, and the remaining four specified that either a pediatrician or an otolaryngologist performed the test. However, whether the test performer was trained or untrained was not specified.

Summary

The meta-analyses revealed that pneumatic otoscopy and professional tympanometry had the highest sensitivity compared with myringotomy. While the diagnostic test with the highest specificity was professional tympanometry (using static compensated acoustic admittance at 0.1), pneumatic otoscopy optimized both sensitivity and specificity. However, the poor quality of many of the studies included in the analysis must be considered. Moreover, most studies failed to provide enough information to assess the qualifications of testers.