Chapter 4. Conclusions

Despite the recent publication of the 1994 Guidelines on Otitis Media with Effusion in Children (Stool, Berg, Berman, et al., 1994), controversy persists over the management and treatment of otitis media with effusion (OME). The present evidence report systematically reviews the recent literature and provides an update to the 1994 guideline on the diagnosis and the late effects of OME on speech, language, and hearing. In addition, while the 1994 OME guideline (Stool, Berg, Berman et al., 1994) did not formally assess OME resolution rates, the present report reviews the natural history of OME. Although the technical expert panel proposed 20 potential key questions for this report, our time constraints allowed us to address only the four key questions judged by the panel to be most pressing and to have the most significant body of recent literature. We did not address any key questions on treatment or management. Of a total of 449 articles originally identified for this report, only a small percentage qualified for inclusion in the synthesis. Thus, the number of meta-analyses performed was a fraction of those we had hoped would inform this report. Our conclusions should be read with this perspective in mind.

Natural History of OME

Based on the available evidence, this report assessed the natural history of OME in terms of resolution rates. The availability of relevant studies limited our analysis to an assessment of natural history at two periods of followup and only for children 3 years of age and older; we were unable to perform meta-analysis for the under-3 age group. For children older than 3 years of age, we were able to conduct two sets of meta-analyses on the resolution of otitis media with effusion at two followup intervals. These sets were matched by unit of analysis, age group, and OME type and, when possible, diagnostic method.

The first set of meta-analyses assessed resolution at 6 weeks followup. In children older than 3 years of age with OME of unknown duration at the onset of the study, 37.2 or 42.3 percent of ears with OME were free of OME at 6 weeks followup, without regard to their interval OME status, depending on the tympanometric diagnostic criteria for OME resolution. Spontaneous resolution rates were significantly different among the cohorts included in the first set of meta-analyses.

The second set of meta-analyses assessed resolution at 3 months followup. Over a period of 3 months, in studies with cumulative resolution rates, and in children older than 3 years of age with OME of unknown duration, 22.5 or 42.7 percent of ears with OME resolve, depending on the tympanometric diagnostic criteria for OME resolution. A disadvantage of using ears as the unit of analysis is the clinical interpretation of the resolution rate: the distribution of OME resolution using ears cannot be assumed to be the same when children are the unit of analysis. Spontaneous resolution rates were not significantly different among the cohorts included in the second set of meta-analyses.

Apart from age, tympanometric criteria for OME diagnosis, and criteria for resolution on followup, which were reported in the studies, we did not have sufficient information from the studies to consider other influential factors. It was also not possible to know with certainty if the children in the various cohorts presented at identical stages in the course of OME. It is unclear why the non-cumulative, 6-week resolution rate estimate should overlap the cumulative,

3-month resolution rate estimate to such an extent. The non-cumulative resolution rate would be expected to include children whose OME had resolved and then re-accumulated middle-ear effusions (which would make the non-cumulative rate greater than the cumulative rate). Therefore, we view these estimates of OME resolution with great caution.

In the Results section, we also described isolated studies of daily followup, resolution of OME of at least 3 months duration at study onset, and a study that derived OME resolution rate equations. The findings of these studies were too limited to draw any broad conclusions. Similarly, a few studies assessed the effects of influencing factors on OME resolution, but, again, we hesitate to draw any generalizations based on such limited evidence.

Literature on the natural history of OME was difficult to interpret for several reasons. These reasons include its generally poor quality of the research, the lack of control for therapeutic interventions, the inability to distinguish persistent as opposed to recurrent OME due to the length of followup, and the varied criteria for continued followup from examination to examination. We recognize that the meaning of quality summary scores, as measured in this analysis by scoring of documentation in the published articles, can be ambiguous (Jüni, Witschi, Bloch et al., 1999). Nevertheless, whether as a summary score or considering individual quality domains, the quality of the twenty-eight cohort studies on natural history was poor. Twenty-four of twenty-seven of these studies had a quality score of three or less on a scale of one (lowest) to six (highest), even though our preselection process guaranteed all cohort studies a minimum score of one. Half of the studies that attempted to assess the natural history of OME did not control for or document control of interventions, either medical or surgical, that might affect OME outcome during the study period. The majority of these investigations did not stratify findings by intervention status. In addition, the intervals between examinations for OME in these studies varied from one day to 3 years; and, it was impossible to assess whether the "continuing" presence of OME, especially after a long interval, represented persistent or recurrent OME. Furthermore, the criteria for followup varied among studies. Most studies continued followup for the whole study period regardless of OME status at a particular exam, but four cohorts terminated followup of individuals who had type A or normal tympanograms at any exam.

Differing definitions of OME resolution and diagnostic methods also make comparison difficult. As noted in the Results chapter, a difference in the definition of OME resolution between tympanogram type B or C transition to A and tympanogram type B transition to A resulted in a three-fold difference in the estimate of OME resolution in one case. As can be seen from our assessment of the operating characteristics of various OME diagnostic methods in the response to Key Question 4 in the Results, the apparent OME resolution rate could be greatly influenced by the sensitivity and specificity of a test.

A paucity of research studies made it impossible to address several other issues. These included the use of the child or the episode as the unit of analysis. Furthermore, we could not assess OME resolution in younger children or the resolution of OME that was not newly diagnosed and of unknown duration.

The literature was lacking in evidence addressing potential factors that might influence OME resolution. A few studies analyzed the potential influence on OME resolution of factors such as gender, at home care versus daycare, season of onset, side of affected ear, race/et`nicity, or diagnostic instrument. Because of the paucity of such studies, quantitative synthesis was not possible. Such information is also necessary to allow generalizations to specific clinical situations.

Similar to the results of Rosenfeld (1999a), our pooled OME resolution rates could be stratified by age and by tympanometric criteria for OME. Although Rosenfeld did not specify his tympanometric criteria, his analysis found a one-month OME resolution rate of 52 percent (95% CI: 47%, 58%) for OME of unspecified duration in cohort studies with ears as the unit of analysis, regardless of age. This finding is similar to our observation of 42.3 percent resolution (95% CI: 24.1%, 60.6%) for children older than 3 years of age. This finding was not based on cumulative resolution rates but on the disappearance of OME at various points in time in children who had OME at the start of the observation period. Further, Rosenfeld (1999a) found a two-to 3-month OME resolution rate of 63 percent (95% CI: 60%, 63%) for OME of unspecified duration in cohort studies with ears as the unit of analysis, regardless of age. This rate is somewhat higher than our estimate of 42.7 percent (95% CI: 29.3%, 56.1%), which was based on studies that measured cumulative resolution.

We restricted our analysis to prospective cohort studies to provide better assessments of the natural history of OME in the general population of children. Nevertheless, we recognize that information on the natural history of OME may also be gleaned from analysis of the results of randomized, controlled trials of treatment of OME, as Rosenfeld (1999a) subsequently reported. Such estimates would be more applicable than would cohort studies on the general population of children to those who, for whatever reason, are being followed closely for their OME by their health provider. Children being screened for OME in the general population might not necessarily have presented to the health care system for evaluation. As a result, these children might represent a less severe class of OME or at least a different population from those children who were already identified with and being followed for OME and who were deemed potential candidates for therapeutic intervention. Rosenfeld (1999a) reported OME resolution rates of 12 percent (95% CI: 8%, 16%), 23 percent (95% CI: 21%, 26%), and 24 percent (95% CI: 17%, 32%), at two weeks, one month, and one to 3 months followup, respectively, in children with OME that had lasted weeks or months and were randomized to the placebo or no-drug arms of randomized controlled trials. In addition, he assessed OME resolution rates of children with OME that lasted 3 months or longer who were surgical candidates for tympanostomy tubes. He found the following OME resolution rates at 6 months, 1 year, 2 years, 3 years, and 4 years for children assigned to the no intervention groups: 27 percent (95% CI: 20%, 35%), 32 percent (95% CI: 24%, 40%), 31 percent (95% CI: 25%, 37%), 49 percent (95% CI: 37%, 62%), and 59 percent (95% CI: 46%, 72%). A strength of these studies is that the diagnosis of OME tended to rely on pneumatic otoscopy as well as tympanometry, rather than on tympanometry alone. A major difficulty of these studies, as with the cohort studies, is the inability to account for interventions received from other providers and beyond the control of the investigators. In fact, one of these investigators mentioned that “Antibiotic treatment was not controlled for in this study” (Maw and Bawden, 1994). Another difficulty, as with the cohort studies, is the inability to distinguish persistent from recurrent OME, due to the long intervals between followup assessments.

Finally, the publication of multiple articles based on the study of a single cohort raises a complex issue. If multiple studies from one cohort were included once in a meta-analysis, the results would be unbiased. However, if findings from a single cohort were included more than once in a meta-analysis, then bias would exist.

Effects of Early-life OM on Long-term Speech and Language Development

Our intent was to examine the influence of long and short-duration OME that results in hearing loss on delays in speech and language development and the ability of other factors to modify the effect of OME. However, we found few studies that addressed these questions. Only 20 of 112 studies (19%) (on 12 cohorts) met our criteria for consideration for this question. These studies further suffered from lack of uniformity with respect to risk factors studied, type of outcome measured, method of measurement, unit of measurement, and age at outcome determination.

The generalizability of our findings also is questionable, since nine of the twelve cohorts primarily included children from specific ethnic/racial groups or particular socioeconomic groups. Five of the six studies included in the three meta-analyses were based on such cohorts. However, we excluded studies on children with specific medical conditions such as craniofacial defects, primary mucosal disorders, immunodeficiencies, and genetic disorders. The available literature did not address the effect of pre-existing speech, language, or other developmental disorders, and only one study focussed on the effect of bilateral persistent middle-ear effusion.

Based on our criteria, only nine of the twenty studies (45 percent) were of acceptable quality. These studies scored five points or more on an eight-point scale.

Our analysis was limited to assessing whether early-onset OME resulted in delays in expressive and receptive language development and cognitive verbal intelligence. For children older than 3 years of age who had a positive history of otitis media during the first 3 years of life, our meta-analyses on long-term expressive language (three studies), receptive language (four studies), and cognitive verbal intelligence (three studies) showed no effect of early otitis media. The 1994 OME guideline (Stool, Berg, Berman et al., 1994) had determined that meta-analysis of the 14 “adequate” studies was not possible, because of the wide variety of measurement tools used for outcome assessment as well as a lack of standardization of data reporting. We were able to identify articles that had appeared since the release of the 1994 OME guidelines.

Nevertheless, our findings on the possible effects of early-life OM on speech and language development are in general agreement with the conclusion of the 1994 OME guideline (Stool, Berg, Berman et al., 1994). They concluded that " rigorous, methodologically sound research does not adequately support or refute the theory that untreated otitis media with effusion results in speech/language delays or deficits." Our findings were also in agreement with those noted in the 1994 OME guideline that "Conflicting findings among studies can be accounted for in several ways: limitations in the research designs, lack of uniformity of test instrument selection, lack of definition of hearing status, and interactions between otitis media with effusion and other risk factors," as well as differences in populations studied. Thus, we caution clinicians not to generalize these findings to children with the underlying chronic medical conditions that were excluded from this study or to those with pre-existing developmental disorders. In addition, generalization to children with persistent bilateral OM may not be valid, since only one of the studies specifically assessed bilateral as opposed to unilateral or bilateral otitis media.

Several ongoing, prospective studies are assessing the effect of early otitis media on long-term speech and language development of children older than 3 years. However, the results of these studies have not yet been published (Paradise, Dollaghan, Campbell et al., 2000; Feldman, Dollaghan, Campbell et al., 1999; Paradise, Rockette, Colborn et al., 1997; Roberts, Burchinal,

Zeisel et al., 1998; Roberts, Burchinal, Jackson et al., 2000). These studies will provide further data on which to base an assessment of the effect of OM on speech and language development.

Effects of Early-life OM on Long-term Hearing

Although the immediate effect of OM on the conductive aspects of hearing is well recognized, the long-term effect of early OM on sensorineural hearing has not been well established (Madell, 1999). Our findings indicate that early OM may have an effect on long-term hearing.

Of the eight cohort studies we analyzed, four reported percent of hearing loss at 6 to 10 years of age. Children with early-life OM have a 2.6 times higher risk of hearing loss at 6 to 10 years of age than do children with no early history of OM with an estimated rate difference of 11 percent. The rate difference and the risk ratio were not significantly different among the cohorts. If early OM does indeed lead to long-term hearing deficits, the clinical implications are significant. Depending on the degree of hearing loss, the hypothesized effect of OM on speech and language development as mediated through hearing loss may be of greater duration than would be explained by the intermittent and transient episodes of conductive hearing loss associated with OM. In addition, the question of the effect of OME treatment on long-term hearing gains greater relevance.

The literature available for assessment of the long-term effects of early-life OM on hearing is both limited and of poor quality. Only four percent (eight of the 186) of the studies that addressed OM and hearing qualified for inclusion in our analysis, and only half of these were of acceptable quality. The evidence on the effects of early OM on long-term hearing also suffered from the same methodological issues as the evidence on the effects of early-life OM on long-term speech and language development. Because of the limited nature of this evidence, and because the rate of intervention depends greatly on the threshold hearing level adopted, the findings of our analysis should be used with caution.

Diagnostic Methods for OME

Previous assessments of diagnostic techniques for OME have recommended several different techniques. The 1994 OME Guidelines recommended pneumatic otoscopy (Stool, Berg, Berman, et al., 1994).

Using 52 diagnostic studies, we were able to evaluate the diagnostic accuracy of the following eight methods: acoustic reflectometry at ≤5 or >5 RU; pneumatic otoscopy; portable tympanometry; professional tympanometry using static compensated acoustic admittance at 0.1, 0.2, and 0.3; professional tympanometry using B curve as abnormal; and professional tympanometry using B or C2 curves as abnormal. All comparisons used myringotomy as the reference standard.

Among the eight diagnostic methods, the receiver-operator characteristic points (plotting sensitivity against [1 minus specificity]) for pneumatic otoscopy were closest to the optimal point of 100% sensitivity and 100% specificity. The pooled sensitivity for pneumatic otoscopy was 94 percent (95% CI: 91%, 96%); the pooled specificity was 80 percent (95% CI: 75%, 86%); the positive predictive value was 89 percent (95% CI: 87%, 92%); the negative predictive value was 89 percent (95% CI: 86%, 93%), and the accuracy was 89 percent (95% CI: 87%,

91%). These findings were based on 2,694 children from seven studies that reported a pooled prevalence rate for OME of 63 percent (95% CI: 58%, 67%). The prevalence rates among the studies ranged from 56 percent to 71 percent, which was a significant variation ($p<0.001$). 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).

Our findings are in general agreement with the recommendations and findings of the 1994 OME guideline (Stool, Berg, Berman et al., 1994). The 1994 OME guideline stated that "The diagnostic evaluation of suspected otitis media with effusion should include pneumatic otoscopy. Otoscopy alone (without the use of the pneumatic otoscope to test tympanic membrane mobility) is not recommended." This recommendation was deemed a strong one based on limited scientific evidence and strong Panel consensus. Based on limited scientific evidence and expert opinion, the 1994 OME guideline also allowed the following option: "Tympanometry may be used as a confirmatory test for otitis media with effusion." The OME Guidelines did not include quantitative syntheses of the evidence.

Our analysis considered several references not cited by the 1994 OME guideline; some were accepted, and some were rejected. The results of our meta-analyses confirm that pneumatic otoscopy had the best operating characteristics among the nine alternatives examined. Our findings also confirm that certain, but not all, categories of tympanometry also perform well in identifying middle-ear effusion in OME as well as in distinguishing it from other entities. While the 1994 OME guideline did not make a recommendation regarding acoustic reflectometry, our findings suggest that acoustic reflectometry does not perform as well as pneumatic otoscopy and certain types of tympanometry. The use of the spectral gradient angle, as the unit of measurement, may improve the sensitivity of acoustic reflectometry compared to the use of reflectivity, but this observation is based on a single study that found a sensitivity of 95.4 percent using a threshold of 95 degrees (Barnett, 1998) (compared to the pooled sensitivity of 64.2 percent for a threshold reflectivity of 5 found in this study. However, the specificity was 31.5 percent when a spectral gradient angle of 95 degrees was used as the threshold, compared to the pooled specificity of 80.4 percent with a threshold reflectivity of 5 in this study. Unlike the 1994 OME guideline, which commented on the combination of tympanometry and pneumatic otoscopy, we did not assess combinational diagnostic methods or algorithms.

The finding that pneumatic otoscopy can do as well as or better than tympanometry and acoustic reflectometry has significant practical implications. For the typical clinician, pneumatic otoscopy should be easier to employ than other diagnostic methods. The important question may be what degree of training will be needed for the clinician to be as effective with pneumatic otoscopy as were the examiners in the studies reviewed in this report. Also, while we did not do a cost-effectiveness analysis, the cost of pneumatic otoscopy, in terms of direct and indirect costs, would appear to be less than that for tympanometry or acoustic reflectometry.

Because of inadequate evidence, we could not conduct evaluations of clinical signs and/or symptoms, air and/or bone threshold audiometry, binocular micro-tympanoscopy, and non-pneumatic otoscopy in the diagnosis of OME. In addition, diagnostic methods that use algorithms or aggregated scorings are important but were not included in the scope of this evidence assessment. The sources of variation of such combinational methods are difficult to detect in published articles. In addition, we must emphasize that we assessed the diagnosis of OME middle-ear effusion at single points in time rather than the diagnosis of persistent or recurrent OME over time. The meta-analyses of diagnostic tools raised several methodological concerns. One concern centered on pooling data from studies of diverse populations. The

differences in OME prevalence among the studies point to one aspect of this diversity. In addition, we were concerned about the different models of instruments within the broad categories of diagnostic methods. We assume instrumentation has improved or at least changed over time in the areas of tympanometry and acoustic reflectometry as well as in the other diagnostic methods that could not be assessed in depth.

In addition, as alluded to above, a definitive assessment of diagnostic methods for the diagnosis of middle-ear effusion in OME would require an assessment of the cost-effectiveness of the different diagnostic procedures. Such an assessment could not be incorporated into our analysis due to a limited time frame. Cost-effectiveness analysis would take into account the specific impact of test results, including false positives and false negatives, which will depend on examiner proficiency and the patient preferences for specific outcomes. Cost-effectiveness analysis would also establish optimal operating points or thresholds for diagnostic methods measured on both ordinal and continuous scales (Sox, Stern, Owen et al, 1989).

Study Limitations

Several limitations of the study applied to the evidence analysis as a whole. These included the issues of the selection of key questions, publication language as an exclusion factor, study quality, and the analysis of influencing factors. This section addresses these issues. Limitations that pertained to individual questions are discussed in the Results section.

Question Selection

Several peer reviewers commented on the absence of OME treatment as a topic of inquiry of this study. The key questions were selected using a standard consensus methodology, as described in the Methods section. The technical experts were asked to consider various factors when selecting questions. One of these factors was the importance of the question (as assessed by the potential impact on OME outcomes and on development of future OME guidelines). Another consideration was feasibility (as assessed by the ability to complete the study in 6 months and the availability of sufficient information or new information, if a systematic review had already been done in the past). Therefore, treatment questions were not arbitrarily included or avoided; instead, adherence to our key question selection method led to the questions that were included. Whether the methodology for selecting key questions needs to be reconsidered in future systematic reviews is a valid question. In addition, whether the composition of a particular technical expert panel might influence the selection of key questions is also a valid question for future systematic reviews.

Study Language

For this analysis, we did not include literature published in other languages, although we did consider studies conducted in non-English-speaking nations but published in English-language journals. This decision was based on our previous experience of limited yield from non-English language publications for our evidence assessment of the management of acute otitis media (Takata, Chan, Shekelle et al., in press; Chan, Takata, Shekelle et al., in press). In that assessment, we reviewed a total of 97 articles published in non-English-language journals and

found two eligible for inclusion. However, these studies were also reported in English language publications. In **Table 55**, we tabulated the country of origin of the 112 studies included in the evidence tables. Of the 108 studies that specified country of origin, 55 (51%) were from (15) countries other than the United States. The percentage of studies from countries other than the United States was 82 percent (27 of the 33) for the natural history question, 50 percent (24 of 48) for the diagnostic methods question, 43 percent (three of seven) for the hearing question, and 5 percent (1 of 19) for the speech and language development question. We also observed that among our peer reviewers, four of whom were European (including two from non-English-speaking countries), none mentioned any specific studies from the non-English-language literature that they believed should have been included in the analysis, based on our study criteria.

Study Quality

For two reasons, no prospective cohort studies were excluded based on study quality. First, we decided that if sufficient numbers of studies were available, we would study the variability of findings by either stratified or sensitivity analyses. Secondly, we recognized the potential problems with summary quality scores (Jüni, Witschi, Bloch et al., 1999). For example, we recognized that a study may have adhered to a high quality of study design but not have documented that design in the article. The abstracted data may be viewed along with our quality score, based on the design described in the article, in the evidence tables.

Influencing Factors

As we mentioned in the Results section, we were unable to conduct in-depth analysis by influencing factors for any of our key questions, due to the constraints of the available evidence. In particular, our assessment of the effect of early-life otitis media on long-term speech, language, and hearing was limited to an assessment of the effect of OME duration. The available evidence was not sufficient to allow us to address the second part of the question, namely, the influence of other risk factors, using standard analytic techniques within the resources and timeframe of this evidence analysis. The technical expert panel had listed many demographic (including SES), environmental, and clinical factors that might either act independently or interact with OME to affect speech and language. To address this question, a meta-regression approach would be required to identify the risk factors that contribute significantly to speech, language, and hearing delays in the context of otitis media and could include both comparative and single cohort studies. However, many issues must be addressed in order to set up data for meta-regression analysis appropriately. Such issues would require a great deal more input from technical experts than was possible with the resources available in this study.