Chapter 2. Methodology

Nomination of Technical Experts

Eleven organizations were contacted for technical expert and peer reviewer nominations. They included the American Academy of Family Physicians (AAFP), the American Academy of Pediatrics (AAP), the American Academy of Otolaryngology-Head and Neck Surgery Foundation (AAO-HNS), the Ambulatory Pediatric Association (APA), the American Academy of Audiology (AAA), the American Speech-Language-Hearing Association (ASHA), the Society for Ear Nose and Throat Advances in Children (SENTAC), the National Association of Pediatric Nurse Associates and Practitioners (NAPNAP), the American Association of Health Plans (AAHP), the Foundation for Accountability (FACCT), and Family Voices**.**

Upon receiving nominations from the agencies, we identified 12 technical experts to serve on the panel. Included were two family physicians, two otolaryngologists, three pediatricians, one audiologist, one speech and hearing expert, one managed-care representative, one nurse practitioner, and one consumer. **Table 1** lists the membership of the Technical Expert Panel.

Topic Assessment and Refinement

A draft work plan for the topic assessment and refinement phase was mailed to the technical experts and representatives of our partners (the American Academy of Pediatrics, the American Academy of Family Practice, and the American Academy of Otolaryngology-Head and Neck Surgery) for review and comments together with a preliminary review that provided a summary of:

- Incidence and prevalence of otitis media with effusion, treatment and management alternatives, the characteristics and size of the affected populations, and the most affected practice settings and providers;

- The burden of illness associated with otitis media with effusion, including morbidity and mortality.

- Extent to which variation exists in practices associated with the prevention, diagnosis, treatment, or diagnosis and treatment of otitis media with effusion.

We reviewed and compiled four previously conducted evidence-based analyses on otitis media with effusion. Particular attention was given to the report of The Otitis Media Guideline Panel on “Managing Otitis Media with Effusion in Young Children” (Stool, Berg, Berman et al., 1994) We distributed these reports to the panel of technical experts for review and for preparation of the initial telephone conference call during which assessment and refinement of the topics wereassessed and refined.

Identification of Key Questions

Prior to the first conference call, we asked each technical expert to submit questions for consideration in the evidence report. The project team organized the responses and compiled an initial list of 20 key questions from the original task order and the letters from the nominating agencies (**Table 2**). This document was distributed to the technical experts after the first conference call, together with a polling form requesting the ranking of the top 10 key questions. The technical experts were asked to rank 10 of the 20 questions from 10 to 1, using 10 as the most important and 1 as the least important. The criteria for ranking were:

1) importance, which included
   a) potential impact on OME outcomes and
   b) b) potential impact on development of future OME guidelines by the partner organizations; and

2) feasibility, which included
   a) possibility of conducting a literature search, review, and data synthesis in 6 months,
   b) availability of sufficient information (data) in the literature to answer the question, and
   c) if applicable, sufficient new information (data) available to affect the results of the last systematic review of the question significantly.

The polling results are tabulated in **Table 3** and the comments are included in **Table 4**. The results were distributed to the experts for discussion during the second conference call.

The four top ranking key questions were selected for consideration in the evidence report. After several revisions at the suggestions of the technical experts, the wordings of the four top ranking (key) questions are as follows:

Key question 1: On Natural History

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?

Key question 2: On Speech and Language Development

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

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

b) What are the risk factors that modulate the effect of OME on speech and language development in infants and preschool children?

Key question 3: On Hearing Decrease

a) 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?

b) 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?

Key question 4: On Diagnostic Methods

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 un-validated examiner
- binocular micro-tympanoscopy
- portable tympanometer
- professional tympanometer
- quantitative tympanometry
- acoustic reflectometry (specify model and year)
- otoacoustic emissions
- audiometry, air or. bone conduction thresholds.

The reference standards to be used in evaluating these diagnostic tests will include tympanocentesis, sedated or non-sedated; MRI; myringotomy, sedated or non-sedated; validated pneumatic otoscopy; and CT Scan.

Identification and Refinement of Causal Pathways, Study Populations, Practice Settings, and Target Audience

The project staff developed a causal pathway and a scope for each of the key questions. We distributed these documents to the panel of technical experts for review and for preparation for the initial telephone conference call. During this call, we assessed and refined the topics and discussed the proposed key questions, target condition, patient populations, clinical context, interventions, and outcomes of interest. The following characteristics of outcomes were proposed: 1) outcomes would be divided into short term and long term; 2) long term outcomes would consist of percent time with effusion, frequency of acute otitis media, hearing loss, speech and language performance, cognition, academic achievement, and other developmental outcomes; 3) duration of short term outcomes would be defined as four weeks or less or eight weeks or less; and 4) duration of long term outcomes would be defined as greater than one year.

During this first conference call, the Technical Expert Panel decided to use the OME Guideline definition of OME: "fluid in the middle ear without signs or symptoms of ear infection". (Stool, Berg, Berman et al., 1994) Though the technical experts agreed on the definition of OME they could not agree on which signs or symptoms should be absent, i.e. what signs or symptoms differentiated OME from acute otitis media. During the first conference call, the technical experts advised us to avoid the use of the terms 'acute', 'subacute', or 'chronic' as descriptors of OME. Instead we should use the duration of OME, such as "under 3 months" versus "greater than or equal to 3 months" as descriptors.

Before the second conference call, the project team developed a draft of the conceptual framework for the proposed key questions. During the second conference call, the framework was discussed and the inputs of the experts were incorporated into the revised overall causal pathway (**Table 5**) from which the causal pathways of the final four key questions were developed. The causal pathways were distributed to the technical experts for further review and comment. The final version causal pathways for the four key questions are presented in **Tables 6 through 9**.

Based on the causal pathways and the discussions during the two conference calls, we developed the scope for each of the four key questions. The scope specifically defined the disease entity, study population, practice settings (including provider type), time period in practice setting, exclusion factors, interventions, influencing factors, outcome measures, literature sources, language, and study design for each key question to be included in the evidence report. We conducted a second poll of the technical experts on each of these domains in which we sought their approval, disapproval, or recommendations for revision on each domain. **Appendix B** presents the version of the scope distributed to the technical experts for polling of their comments and approval. **Appendix C** presents the results of the polling of the experts' comments on the scope. In response to these comments, we further revised the key questions, causal pathways, and the various domains of the scope. We incorporated comments from technical experts to the extent possible, except those related to other domains, those that were obvious misunderstandings or misinterpretations, or those suggestions for deletions or additions that could be handled during the analysis phase of the project. Specifically, the project

team retained several influencing factors recommended for deletion by some experts. The project team took note of the deletions and would later stratify the analysis, if possible, by the group of factors unanimously recommended by the experts versus those that were not. Such stratified analysis would depend on the number of studies that specifically address these factors. We revised the key questions according to experts' suggestions. We further revised the scope and reworded the key questions according to the final round of comments: the final version of the scope is included in **Appendix D**.

In preparation for supplemental analysis, we took polls to solicit the technical experts' opinion on the importance of the risk factors identified in the analytical framework. Specifically, we asked the experts the following questions for each factor:

a) Regarding Key Question 1: "Does this factor influence the natural history of OME?"

b) Regarding Key Question 2: "Does this factor have an independent effect on speech and language development separate from its effects on OME or unspecified OM?"

c) Regarding Key Question 3: "Does this factor have an independent effect on long-term hearing separate from its effects on OME or unspecified OM?"

d) Regarding Key Question 4: "Does this factor have an independent effect on the accuracy of a diagnostic method separate from its effects on OME or unspecified OM?"

The experts had a choice of responding "yes", "no", or "don't know." For each such opinion, the experts were asked to indicate the basis of their opinion by choosing one or more of the following: "Judgment/Experience," "Theoretical Construct," or "Literature".

The questionnaire for the two polls is included in **Appendix E** and the responses of the 12 technical experts are presented in **Appendix F.** A summary of the risk factors ranked by the importance assigned by the technical experts is presented in **Table 10**.

Literature Search

The Technical Expert Panel and project staff developed a literature search strategy. The literature search included the search of three databases: MEDLINE (1966-January2000), the Cochrane Library (through January 2000), and EMBASE (1980-January 2000). We identified additional articles by review of reference lists in proceedings, published articles, reports, and guidelines.

The project librarian developed an overall search strategy for MEDLINE (**Appendix G**) that incorporated the input from the technical experts and followed the scope of the project. The MEDLINE database is produced by the U.S. National Library of Medicine and is widely recognized as the premier source for bibliographic coverage of biomedical literature. It encompasses information from Index Medicus, Index to Dental Literature, and International Nursing, as well as other sources of coverage in the areas of allied health, biological and physical sciences, humanities, and information science as they relate to medicine and health care, communication disorders, population biology, and reproductive biology. We searched the MEDLINE database for publications dating back to 1966. Further, we included articles in the English language only for the following reasons. First, our experience with our evidence assessment of the management of acute otitis media (Takata, Chan, Shekelle et al., in press;

Chan, Takata, Shekelle et al., in press) demonstrated a low yield from non-English language publications (only two studies accepted out of 97 reviewed and both of theses were also published in English). Second, we needed to balance limited resources between reviewing non-English language literature and answering additional key questions. Since empiric evidence of the need to include non-English language literature in meta-analyses was mixed (Moher, Pham, Klassen et al., 2000) and reviewing non-English literature would be resource intensive, we chose to limit our scope to English language literature only.

The MEDLINE search strategy used both controlled vocabulary terms and keywords. The strategy was organized into modules or clusters of search statements. The main groupings included: otitis media with effusion (OME), mastoid, otitis media; natural history; speech and language; hearing; and diagnosis. These groupings corresponded to the key questions.

For the "otitis media with effusion" concept, both the controlled vocabulary term otitis media with effusion and text word were used. A variety of additional terms were used; such as allergic otitis media, fluid ear, glue ear, middle ear effusion, mucoid otitis media, nonsuppurative otitis media, secretory otitis media, and serous otitis media. For the "mastoid" concept, both the controlled vocabulary and the text word were used. The otitis media module included what is referred to as an "explode" of otitis media, which included the controlled vocabulary headings "otitis media", "otitis media with effusion", and "otitis media, suppurative."

The "natural history" module combined "OME" or "mastoid" with a combination of text words and controlled vocabulary terms for natural history including "natural course", "natural history", "placebo", "placebos", "resolution", "self limited", "self limiting", "untreated", and a variety of terms for spontaneous resolution.

Both the "speech and language" module and the "hearing" module combined OME, or mastoid, or an explode of "otitis media" with the speech, language, and hearing concepts. The speech and language component used the controlled vocabulary terms for speech and language, speech and language disorders, child language, communication, communication disorders, language development and tests, voice, and voice disorders. In addition, the text words "speech" and "language" were added. The hearing module used the controlled vocabulary terms for hearing and hearing disorders and hearing aids and tests, as well as the text word "hearing".

The "diagnosis" module combined OME or mastoid with a combination of text words and controlled vocabulary terms for diagnosis. In addition to the controlled vocabulary terms for diagnosis and diagnostic techniques and procedures, a number of text words were added for audiometry, diagnosis, diagnostic, otoscopy, and tympanometry.

We customized the search strategy initially developed for MEDLINE for EMBASE. EMBASE, the Excerpta Medica database, produced by Elsevier Science, is a major biomedical and pharmaceutical database that indexes over 3,800 international journals. EMBASE is one of the most widely used biomedical and pharmaceutical databases. The database currently contains over 6 million records, with more than 400,000 citations and abstracts added yearly. We searched the EMBASE database for citations dating back to 1980. For the search in the Cochrane Library, we used "otitis media" as the search term. The Cochrane Library contains several databases: (1) The Cochrane Database of Systematic Reviews, which contains Cochrane reviews published by the Cochrane Collaboration, an international organization dedicated to applying evidence-based-medicine principles to the review of important clinical topics; (2) The Cochrane Controlled Trials Register which is a bibliographic database of controlled trials; (3) The Database of Abstracts of Reviews of Effectiveness (DARE), which includes structured abstracts of systematic reviews that have been critically appraised by reviewers at the NHS Centre for Reviews and Dissemination in York and by other experts, for example from the American College of Physicians' Journal Club and the journal Evidence-Based Medicine; and (4)

The Cochrane Review Methodology Database which is a bibliography of articles on the science of research synthesis.

The Cochrane Library search yielded 666 titles/abstracts. The MEDLINE search resulted in 2,379 titles/abstracts. After eliminating duplicates, we retained 2,207 titles/abstracts for screening. The EMBASE search retrieved 1980 citations. After eliminating duplicates, we retained 327 for screening.

We conducted all searches in January 2000 and subjected a total of 3,200 titles/abstracts to screening by two physician reviewers. By merging and eliminating duplicates from the titles/abstracts from the three databases, we created a database of titles and abstracts.

EndNote software (EndNote Windows Version 3.0, 1st Edition. Niles Software Inc., Berkeley, CA) was used to keep a complete record of all titles/abstracts and identify duplications. This software stores, organizes, and tracks references by source (e.g. identified in MEDLINE), search strategy (date of search, index code specifying search criteria used), and a unique identification (UI) code for each article (assigned by the source used to find article). Electronic removal of duplicate citations was supplemented by manual cross-checking. In the event an article was identified by an expert panel member or through reference checking, the title and author of the reference were entered into MEDLINE through the Ovid search system (Ovid Technologies, Inc. 1998, Version: 7.8 Millennium source ID 1.3932.1.156.1.7, Revision: 1.303.2.8) to determine the UI. If a UI could not be found for the article, an alternate identification code was assigned.

EndNote assigned a record number to each new reference added to the master file. This number would not change once an article was added to the list and was used, in addition to the UI, to sort references for article retrieval and review.

Upon completing the literature search and duplicate checking, we exported the master list generated from EndNote to a Microsoft Excel spreadsheet for data export and analysis. We added codes including status of article retrieval, reviewer, and the results of the review.

Review of Retrieved Titles/Abstracts Against Screening Criteria

After retrieving of titles and abstracts from the literature search, two physician reviewers reviewed the abstracts against the inclusion/exclusion criteria to determine eligibility for inclusion in the evidence synthesis as defined in the scope and key questions. Titles/abstracts were not masked prior to review. A pre-designed screening form was used to record the reviews. A meeting was held to review the instructions for screening (**Appendix H**), including the use of the computerized data forms. The reviewers entered the screening results directly into the computer and forwarded the results electronically to our data analyst for processing. The screening results for each title/abstract were matched between the two reviewers, and discrepancies on inclusion or exclusion were resolved in conference calls among the two reviewers and the task order coordinator. The data analyst generated summary reports indicating those abstracts that passed the screening criteria and those that failed and the reasons for failure.

We completed screening of the 3,200 titles/abstracts from the Cochrane Library, MEDLINE, and EMBASE. After resolution of 376 discrepant citations, 2230 (70%) were rejected and 970 were accepted for full article review. The reasons for rejection of the 2230 citations are presented in **Table 11**. We also screened the database provided to us by the American Academy of Pediatrics (AAP) from its recent review of the topic. Of a total of 1918 titles/abstracts screened against our database, we identified 477 duplicates from our ENDNOTE database, leaving 1441 records from the AAP files that required further screening by the two reviewers.

The screening of the 1441 citations from AAP files identified 32 additional citations that required full article review.

The third source of reference material was the six proceedings of the International Symposium on Recent Advances in Otitis Media with Effusion from which we identified 159 additional citations for full article review.

The fourth source consisted of references in books and articles from which we identified 31 full articles for further review.

Retrieval and Review of Full Articles

The titles/abstracts identified for further review were forwarded to the library for full article retrieval. Libraries at both the Los Angeles County - University of Southern California Medical Center and the University of Southern California Health Sciences Campus were the primary sources of the articles. Those not found were retrieved through the Inter-Library Loan Program.

Because a large number of titles/abstracts had inadequate information for full evaluation, a secondary screening of full articles was conducted. Two physicians or a physician and a health services researcher reviewed each article. Articles were not masked prior to review. Discrepancies on inclusion/exclusion were resolved between the reviewers.

Of the 1,250 full length articles to be retrieved, 3 were irretrievable due to incorrect citation information. Secondary review of the remaining 1,247 full length articles from the various sources resulted in the rejection of 798 articles and acceptance of 449 articles: 141 for question 1, 112 for question 2, 186 for question 3 and 75 for question 4. **Table 12** provides the reasons for rejection of the 798 articles and **Table 13** summarizes the number of articles accepted during the secondary review process. During the fourth conference call, the experts raised the age limit to 22 for the responses to questions 2 and 3 to allow for detection of speech, language, and hearing problems past age 12, the original upper age limit. As a result, we revisited all titles/abstracts and articles that had been rejected because of the age limit, and four previously rejected articles were accepted from our original databases.

After establishing the analytical plan and before data abstraction, a physician and one health services researcher carried out a tertiary review of the 449 articles according to a set of established criteria for each key question. During this tertiary review, study design and quality were also evaluated.

For Question 1, we used three criteria for tertiary screening: 1) was the study a prospective cohort(s) study or a randomized control trial (RCT), 2) whether the control group in the RCT used the other ear as the 'unit of control' or not, and 3) were the outcome data abstractable?

For Questions 2 and 3, we established 5 criteria for tertiary screening:

1) Was the degree of OME determined for the first 3 years of life, and could the OM degree for the period before 3 years of age be linked to a specific outcome? If a study began at age 3, the study was not considered to fulfill this criterion.

2) Was the upper age limit 22 years of age? If a study included subjects older than 22 years, this criterion could be fulfilled only if outcomes for the 22 years of age and under was reported.

3) Was the degree of OM graded in some way, such as total time with OM, some measure of OM persistence, OM recurrence, or some measure of OM severity, and could the OM degree grade be linked to a specific outcome?

4) Was the study prospective? A study is considered prospective if the outcomes were measured prospectively. Cross-sectional and case-control studies were specifically excluded. A study that followed subjects prospectively for both OM history and outcome measures was considered a prospective cohort study. Studies that followed subjects prospectively for outcome measures (i.e. over a period of time) but retrospectively for OM history were considered to be retrospective-prospective studies and were accepted for inclusion. Studies that collected data on the outcome measures at one point in time, were considered retrospective cohort studies regardless of whether OM history was collected prospectively or retrospectively and were excluded. Studies that presented a cross-sectional analysis of prospectively collected data were not considered prospective and were excluded. Randomized controlled studies of an intervention with longitudinally measured outcomes were considered prospective and were included.

5) Was the outcome measured when the child was older than 3 years of age?

For Question 4, we established 3 criteria for inclusion during tertiary review: 1) Were the diagnostic procedure of interest and the reference standard performed within 24 hours of each other? 2) Is the diagnostic procedure not an algorithm or combination of multiple diagnostic procedures, 3) Are the reference standards one of those specified in the scope (tympanocentesis, MRI, myringotomy, validated pneumatic otoscopy, or CT scan), and 4) Are the data abstractable.

After tertiary review and data abstraction, a total of 114 articles were included in this evidence report, five of which addressed more than one key question:

- 38 cohort studies for question 1, the natural history question;
- 21 studies for question 2, the speech and language question;
- 8 studies for question 3, the hearing question; and
- 52 studies for question 4, the diagnostic method question.

Table 14 presents the results of tertiary review of the 449 articles.

Review and Assessment of Study Quality

We established criteria used for the assessment of study quality prior to the review of articles. Only prospective cohort studies were reviewed for Questions 1, 2 and 3 because of concerns about the validity of case-control, cross-sectional, and retrospective cohort studies. Diagnostic studies were reviewed for Question 4. The criteria used to evaluate the quality of

both types of studies were modified from the work by the McMaster University Group (Jaeschke, Guyatt, and Sackett, 1994; Sackett, 1981; Trout, 1981; Tugwell, 1981).

The quality of natural history studies was evaluated against the following criteria:

1) Was the study a prospective cohort study?

2) Was the outcome(s) of the study clearly defined?

3) Was the outcome(s) measured at a clearly defined timepoint(s)?

4) Was the cohort of subjects followed without any intervention?

6) Was there blinded assessment of the outcome(s) of the study?

7) Were point estimates and measures of variability provided for the main adverse outcomes measured?

The quality of prospective cohort studies was evaluated against eight components:

1) Was the study cohort(s) clearly defined, with clearly spelled out inclusion and exclusion criteria?

2) Was the study cohort(s) assembled at a uniform point in the course of the child's illness?

3) Were the pathways by which patients entered the study clearly described?

4) Was complete follow-up achieved?

5) Were withdrawals and drop-outs described?

6) Were objective outcome criteria developed and used?

7) Was the outcome assessment "blind"?

8) Was adjustment for extraneous factors carried out?

The quality of diagnostic studies was evaluated against six components:

1) Was the reference standard appropriate?
2) Were the test results and the reference standard assessed independently of each other?

3) Were the readers of the results of the diagnostic test or the reference standard blinded?

4) Did the patient sample include an appropriate spectrum of mild and severe, treated and untreated patients to whom the diagnostic tests were applied in clinical practice?

5) Were the reproducibility of the test result (precision) and its interpretation (observer variation) determined?

6) Were the methods for performing the test described in sufficient detail to permit replication?

Articles were not masked prior to review. The Task Order Coordinator resolved minor discrepancies between the two reviews of each article. Conferences were held to resolve discrepancies whenever needed.

Data Abstraction

For the articles deemed eligible for inclusion in the Evidence Report, data abstraction was carried out by a two-member team that consisted of a physician reviewer and a health services researcher. One of the two members abstracted the data onto the evidence table, and the other member checked the data for accuracy. Data abstracted included parameters necessary to define study groups, inclusion/exclusion criteria, influencing factors, and outcome measures to be used in analysis.

Specific instructions for data abstraction were recorded. For Question 1, the outcome indicators for abstraction included partial OME resolution (resolution in one ear for bilateral OME only), complete OME resolution, relapse/recurrence (fluctuation/dynamic course), AOM after OME. For Question 2, the outcome indicators for abstraction were expressive or receptive language, expressive or receptive speech, and cognitive verbal intelligence. For Question 3, the outcome indicators for abstraction included conductive or sensorineural hearing loss. For Question 4, the outcome indicators for abstraction were sensitivity, specificity, positive and negative predictive values, and prevalence rate. For all questions, the time or age at which each outcome was measured was recorded. The outcome measures included both continuous and categorical measures. Continuous measurements included mean time or a median time and the categorical measures included proportion with resolution at specified times where both numerators and denominators were recorded. A key issue was that individual children, not populations, must be tracked. The latter would be acceptable only if we knew for certain that the same children were checked at both times.

Procedures to Reduce Bias, Enhance Consistency, and Check Accuracy

To reduce selection bias, we assigned two physician reviewers to screen and review titles/abstracts and full articles at every stage of the selection process. We assigned one physician and one health services researcher who was familiar with experimental design and biostatistics to abstract data. We assessed completeness of our collection of retrieved articles by cross-checking with studies included in other meta-analyses and references listed in review articles. The software program EndNote was used to check batches of articles added to the master list for duplicate references by comparing author, year title, and reference type. Following the importation of the first literature search, we used the software program EndNote to check subsequent references for duplication prior to their addition to the master list. After the master list was completed, we performed a second, manual, check to ensure no duplication.

To assess the extent of publication bias, we searched multiple sources and unpublished material identified by the Technical Expert Panel and internal content experts. We also studied

funnel plots—scatter plots of sample size versus the estimated effect size from each study. When publication bias existed, a portion of points would be missing from the funnel plot, typically at the null effect level. Because graphical evaluation can be subjective, we also conducted an adjusted rank correlation test (Begg, 1999) and a regression asymmetry test (Egger, Smith, Schneider et al., 1997) as formal statistical tests for publication bias. We conducted these tests using the statistical package Stata (StataCorp. 1999).

The mechanisms used to enhance consistency in screening and data abstraction include the use of pre-designed forms with explicit instructions and continuous and prompt resolution of discrepancies. Data were entered into a Microsoft Excel spreadsheet directly by the screeners or data abstractors. A third project staff cross-checked data for individual studies abstracted by each data collector. We resolved discrepancies by rechecking the article or by consensus via conference calls.

Preparation of Evidence Tables

An evidence table was prepared for each key question. Each evidence table provides a comprehensive tabular display of data abstracted from the literature in response to the question. It contains the name of the first author, year of publication, study design and quality score, how and by whom OME diagnosis was done, when and where the study took place, inclusion and exclusion criteria, important influencing factors, sample size, outcome measures and their definitions, and study findings. A total of four evidence tables was prepared; they are included in the section called Evidence Tables in this report.

Supplemental Analysis

Based on the discussions of technical experts during the conference calls and the designated time frame for the evidence assessment, a supplemental analysis plan was developed for each key question to synthesize the data.

Natural History

Question 1. What is the natural history (spontaneous resolution rate over time without treatment) for the following diagnostic groups: a) OME persisting after a discrete episode of acute otitis media, b) newly diagnosed OME of unknown duration (unilateral or bilateral), c) OME persisting for weeks or months (unilateral or bilateral), d) unilateral OME lasting 3 months or longer, and e) bilateral OME lasting 3 months or longer?

The outcome measures for this questions included complete and/or partial resolution rates, relapse/recurrence rates, and incidence of AOM after OME. The scope listed 31 non-treatment factors that might affect the course of the illness and confound the outcomes. They included: age, gender, ethnicity/race, socioeconomic status, number of hours in child care center, tobacco smoke exposure, season, number of children in household, breast-fed status, barotrauma challenges, OME laterality, hearing level, total duration of OME, age at first OM onset age of previous OME, number of previous OMEs, family history of OME, otitis prone (AOM), allergies, prior tubes, prior adenoidectomy, developmental delay, caregiver preference for treatment, caregiver education, examiner skill, examiner type, health care setting, monitoring time, monitoring frequency, monitoring personnel, and monitoring method.

Furthermore, the type of study is an important consideration for the assessment of natural history. A stratified random sample of a broad, well-defined population provides evidence of good generalizability, but may be restricted in the amount of clinical information on participants. A single (untreated) arm from a clinical trial will usually provide much more clinical evidence about OME, but this is usually assessed on a very selected group of children, making generalizing the results to the general population more difficult. For this evidence assessment we used only prospective cohort studies as these came closest to the ideal of enrolling a sample from a broad poulation.

The first step of the analysis was to obtain a distribution of studies stratified by the 5 diagnostic groups (namely, 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), by type of outcome measures, and by non-treatment factors. This stratification provided us with an overview of the emphasis of past research in this area and an opportunity to identify gaps and areas for future research.

Using the DerSimonian and Laird random effects model (DerSimonian and Laird 1986) to pool rates across studies, we performed a meta-analysis on strata with more than 3 studies for a pooled estimate of an outcome with 95% confidence intervals. This method produces a summary measure by weighting each study's measure by the inverse of the sum of the within-study variance and the between-study variance. This approach allowed both sampling variation and between-study heterogeneity to affect the pooled estimate.

In addition to the pooled estimate, we reported the Q statistic and p-value for the Chi-squared test of heterogeneity, which tests the null hypothesis that the individual study results are homogeneous (Laird and Mosteller 1990).

Effects of Early-Life OM on Long-Term Speech and Language Development

Question 2. Do infants and preschool children with longer duration of 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? One specific formulation of this question is: 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?

For the first part of the question, the outcome of interest was speech and language developmental delay and the risk factor of interest was OME-associated conductive hearing loss and/or long versus short duration of early-life OME. For this question we included only comparative studies. Further, since prospective comparative cohort studies provide better evidence than retrospective comparative cohort studies, we conducted our assessment using only prospective comparative cohort studies.

The risk factor of interest was whether a child had or did not have OME-associated conductive hearing loss in the first 3 years of life, or whether duration of OME during the first three years of life was long or short. For this risk factor, we collected data on five related variables: hearing level, total duration of OME≥3 months, number of previous OMEs, duration of middle ear effusion (MEE), and repeated or persistent versus infrequent early-life OME. The hearing level was used to determine whether a child had OME-associated conductive hearing loss in the first 3 years of life. The total duration of OME greater or equal to 3 months was used

to define length of duration. We used the repeated or persistent versus infrequent early-life OME to define the risk. If a study did not classify the study subjects this way and (instead) reported data by number of previous OMEs and/or duration of MEE's, we sought the advice of the technical experts to stratify the samples based on these variables.

The influencing factors of outcome included both treatment and non-treatment factors. Here we are using "influencing factors" as a general term including risk factors for OM and/or confounding factors for the dependent variables of interest. The non-treatment factors included: age at first OM, gender, ethnicity/race, socioeconomic status, number of hours at a child care center, quality of child care, early intervention program, tobacco smoke exposure, number of children in household, breast-feeding status, OME laterality, allergies, developmental delay, OM complications, e.g. perforated TM, cholesteatoma, chronic illness of any type, caregiver education, quality of parent-child interaction, examiner skill, examiner type, health care setting, age at rechecks, frequency of rechecks, primary care provider, and type of equipment to measure hearing. Treatment factors included any combination of the following: tympanostomy tubes, adenoidectomy, myringotomy, antibiotics, systemic steroids, decongestant, antihistamine, N-acetyl-cysteine or others.

The outcome measures for this question related to speech and language developmental delay. These outcomes were measured by different instruments at different times, by different professionals, in different settings. In preparation for information synthesis, with the assistance of our speech and language technical expert, we classified the tests used in our final set of studies into the five outcome categories: expressive language, receptive language, expressive speech, receptive speech, and cognitive verbal intelligence. For analysis, we first stratified studies by the type of outcome measures, risk factor measures, and treatment or non-treatment risk factors. For any comparison among 3 or more studies, we conducted a meta-analysis. In each meta-analysis, we derived a pooled effect size defined as the proportion of standardized difference between the positive and negative otitis media groups. We pooled across studies the standardized mean differences between the groups and divided by a pooled standard deviation. We used a random effects model (DerSimonian and Laird, 1986) and the Hedges estimate of the pooled standard deviation (Hedges and Olkin, 1985). We used Stata (StatCorp. *Stata Statistical Software: Release 6.0.* College Station, TX: Stata Corporation. 1999) for the analyses.

To answer the second part of the question: "What are the risk factors that interact with the effect of OME on speech and language development in infants and preschool children?," we planned to conduct meta-regression analysis to identify the risk factors that contribute significantly to speech and language delays. For this analysis we would include both comparative and single cohort studies. Many technical issues must be addressed to set up data appropriately for meta-regression analysis. Due to restriction of the time frame, this part of the question was not included in this assessment but should be an area of future research.

Effects of Early-Life OM on Long-Term Hearing

Question 3. 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 the following: 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 risk factors interact with the effect of OME on hearing loss later in life (unilateral or bilateral) in infants and preschool children?

The analysis plan for Question 3 followed that for Question 2. In preparation for information synthesis, we sought the assistance of our audiology technical expert, to group the tests used in our final set of studies into homogeneous categories.

Our audiology expert advised that (1) an acoustic reflex at 500 or 1000 Hz were both acceptable for study, (2) the criterion for abnormal reflex threshold would depend on the study, whether ipsilateral or contralateral, and on the frequency used for testing, (3) the abnormal reflex criteria should be based on normative data, and (4) quantitative tympanometry should be classified as:

a. Static Compensated Acoustic Admittance including: peak admittance, peak compensated admittance, peak compliance, static compliance, static admittance, and peak compliance.

b. Tympanometric Gradient including: gradient, pressure gradient and tympanometric gradient (Madsen compliance was excluded because Madsen compliance units were arbitrary units and the this instrument was from an era in which the units were not on a calibrated scale.)

c. Tympanometric Width referring to terms containing the words width, referring to tympanometry.

Diagnostic Methods for OME

Question 4. What are the sensitivity, specificity, and predictive values for alternative methods of diagnosing OME compared with one of the reference standards?

The diagnostic methods to be assessed included: a) signs/symptoms, b) non-pneumatic otoscopy, c) pneumatic otoscopy, validated or unvalidated examiner, d) binaural (or bilateral) micro-tympanoscopy, e) portable tympanometer, f) professional tympanometer, g) quantitative tympanometry, h) acoustic reflectometry, i) otoacoustic emissions, and j) audiometry, air or. bone conduction thresholds. The reference standards used to evaluate the accuracy of the diagnostic methods included: a) tympanocentesis, sedated or non-sedated, b) MRI, c) myringotomy, sedated or non-sedated, d) validated pneumatic otoscopy, and e) CT Scan.

Diagnostic methods based on algorithms, combinations of methods, or combination of scores, were not within the scope of this report because the sources of variation of such combinational methods would be difficult to detect in published articles and the analysis of them would not be feasible within our timeframe. Also excluded were studies where the experimental diagnostic test and the reference standard test were performed more than 24 hours apart.

Our strategy for evaluating the diagnostic value of a procedure was to derive pooled estimates for sensitivity, specificity, and prevalence rate for each diagnostic procedure and reference standard with 3 or more comparison studies. We used the DerSimonian and Laird random effects model (DerSimonian and Laird 1986) to derive random effects estimates and 95% confidence intervals. We also pooled the prevalence rates to determine the heterogeneity of the study populations. Using the pooled estimates, we plotted the performance of each diagnostic test in terms of sensitivity and (1-specificity) and identified the best performer among the tests included in the comparison. We then derived the positive and negative predictive values for the best diagnostic test for various prevalence levels.

To prepare for a meta-analysis for each comparison, we abstracted data from the evidence table; one meta-analysis was performed for sensitivity and specificity,. The following data

elements were entered into the SAS program to be converted into a SAS data set: study ID number, author and year of publication, number of adverse outcomes in the experimental group, total number of patients in the experimental group, number of adverse outcomes in the control group, and total number of patients in the control group. We used a SAS macro software program developed by RAND statistical staff to perform all meta-analyses and used the beta-test version of the software package "MetaGraphs" (1998, Belmont Research, Inc. 84 Sherman Street, Cambridge, MA 02140) for graphing.

The following statistics were generated from the SAS macro program: (a) study-level statistics (incidence rate, relative risk, risk difference, number needed to treat (NNT), odds ratio, and their 95 percent confidence intervals); (b) crude estimates and their 95 percent confidence intervals for all studies combined; (c) fixed effects estimates and their 95 percent confidence intervals for all studies combined; (d) random effects estimates and their 95 percent confidence intervals based on the DerSimonian and Laird method for pooling study results, and Chi-squared test of homogeneity; (e) weight for each study for both the fixed effects model and random effects model used to calculate of risk difference and relative risk.

To use MetaGraph for graphing, we entered the data into ASCII files using the UltraEdit-32 software. Funnel plots were produced for the purpose of screening possible publication bias, and the shrinkage plots were generated to display the effect size of each study and compare it against the overall model estimate, together with the 95 percent confidence limits. We evaluated the funnel plots graphically for asymmetry that resulted from the non-publication of small, negative studies. Because graphical evaluation can be subjective, we also conducted an adjusted rank correlation test (Begg, 1999) and a regression asymmetry test (Egger, Smith, Schneider et al., 1997) as formal statistical tests for publication bias. We conducted these tests in the statistical package Stata (StataCorp. 1999).

Identification of Peer Reviewers

At the beginning of the project, we requested nominations for technical experts and peer reviewers from 12 organizations. A total of 18 nominations were received for the Peer Review Panel. Experts in systematic reviews and meta-analysis were selected from a pool of experts associated with the Southern California Evidence-Based Practice Center but not involved with this project. The Project Staff, in consultation with the Task Order Officer, determined the relative mix of reviewers across the three domains (methodology, user, and clinical). In addition to domestic experts, we identified four European experts to serve as peer reviewers. The Peer Review Panel (**Table 15**) was composed of 18 members including family physicians, pediatricians, otolaryngologists, audiologists, speech-language pathologists, nurse practitioners, health planners, consumers, systematic review methodologists, statisticians, and non-U.S. experts in otitis media.

Peer Review Process

A copy of the draft evidence report was mailed to each peer reviewer on the panel, along with an instruction sheet (**Table 16**) for reviewing the draft evidence report. The Peer Review Panel was asked to respond within three weeks. Seventeen of the 18 peer reviewers responded with comments. A copy of the draft evidence report was also mailed to the members of the Technical Expert Panel and all technical experts responded with comments. Upon receipt of all responses from the peer reviewers and technical experts, the project staff compiled a summary of

the comments and changes and revised the draft evidence report. We forwarded all comments to the Task Order Officer for review. The peer reviewers' and technical experts' comments are included in **Appendix I,** together with the corresponding responses or actions taken by project staff.