Chapter 2. Methods

Overview

This chapter documents the procedures that the RTI-University of North Carolina at Chapel Hill Evidence-based Practice Center (RTI-UNC EPC) used to develop a comprehensive evidence report on the effectiveness over time of adjunctive therapies used in addition to scaling and root planing (SRP) in treating adults with chronic periodontitis. To set the framework for the review, we discuss first the key question that we address and the related underlying causal pathway for this topic. A detailed description of the literature search process follows; it includes descriptions of the Medical Subject Headings (MeSH terms) used in the principal search, other search sources, the inclusion and exclusion criteria, and the application of these criteria to the results of the searches. We note steps for reviewing studies that met the inclusion and exclusion criteria, abstracting data onto data abstraction forms, creating evidence tables, writing a draft report for external peer review, and revising the draft following peer review. We also discuss quality issues, in particular, the RTI-UNC EPC's quality control procedures for determining eligibility for inclusion, carrying out data abstraction, checking evidence and text tables against articles, and grading the quality of individual studies.

Key Question and Casual Pathway

The overarching key question is: How does the effectiveness of SRP therapy by itself for the treatment of chronic periodontitis compare to SRP accompanied by adjunctive therapy at varying lengths of time after treatment? Figure 2, the basic causal pathway for this key question, indicates the characteristics of the patient population, the nature of the SRP services, the range of adjunctive therapies considered, the possible outcomes of interest, and the various time frames in which outcomes might be measured.

We constrained the review to chronic periodontitis among adults and excluded studies pertaining solely to more aggressive forms of periodontal disease. Thus, we did not examine studies relating to treatment of periodontitis that is described as refractory, localized, juvenile, aggressive, related to human immunodeficiency virus or acquired immune deficiency syndrome (HIV/AIDS), or related to diabetes; neither did we consider studies that included a mix of periodontal diseases but that did not present separate analyses of the chronic form of the disease.

Ultimately, we focused on the following adjunctive therapies, which in some cases include both systemic and local delivery modalities:

- Tetracycline
- Minocycline
- Metronidazole
- Metronidazole in combination with amoxicillin
- Chlorhexidine
- Other antibiotics (e.g., spiramycin, doxycycline, azithromycin, Augmentin), and
- Other antimicrobials (e.g., povidone iodine, hydrogen peroxide, fluoride).

The key question defines the population of interest as adults receiving SRP for chronic periodontitis. The principal audience for this review comprises dental practitioners and

researchers, especially in periodontology; however, we believe the issues may be of interest to physicians, and so we did not, in the searches, exclude studies done in medical care settings.

Of all the potential outcome measures in this literature, we had to constrain those for this systematic review to a very small number. As discussed in Chapter 1, after discussion with clinical experts asssisting the team and the Technical Expert Advisory Group, we focused our analyses on two clinical measures: probing depth (PD, measured in millimeters of reduction) and clinical attachment level (CAL, measured in millimeters of gain). Even with this narrow focus, the nature of these measures can vary significantly across studies (i.e., across dental examiners) on the basis of several factors: the level of gingival inflammation at the time of measurement, diameter of the probe tip, probing force, and angulation of the probe,[28] often these factors reflect the nature of the original or graduate training of the examiners, but in any case they raise the level of incomparability across studies to some unknown degree. We also targeted one microbiological measure - presence or percentage reduction in spirochetes - for the qualitative analyses, but we did not try to use this for the meta-analyses, as it was even less often reported than either PD or CAL.

We reported the longest period of follow-up for entering information into evidence tables; when data were provided only by subgroups (e.g., subgroups defined by different levels of PD at baseline), we attempted to retain all that information in evidence tables or, if the volume of information was too large, to highlight data from the subgroups with the worst baseline periodontal disease severity. In some cases, articles otherwise of interest did not report on these measures in any direct way but may have given other empirical evidence, and when such information showed a statistically significant *net* difference between the treatment and control groups (i.e., a difference of the differences between baseline and the end point), we tried to reflect that in discussing results.

Literature Search and Analysis Strategy

Literature Searches

To comprehend fully the scope of chronic periodontitis research, we had to design and implement several different search approaches. We searched for articles on clinical trial research on periodontitis that gave particular attention to SRP. We also searched for specific antibiotics and other chemical antimicrobials to see if those terms produced new articles or only ones we had already found. To find relevant articles on all aspects of periodontitis, we searched MEDLINE for papers published from 1966 through the second week of February 2002. This search is fully documented with the terms (MeSH and key words) and the counts of articles identified at each step of the search process in Table 2. Table 3 provides additional specifications of our inclusion and exclusion criteria.

Our initial MEDLINE search identified 10,670 articles (Table 3). After limiting to trials and human studies, 2,987 articles remained (Step 6). We further searched for “dental scaling” and “root planing” trials, yielding 836 articles, then searched for the specific drugs listed in Table 2. We had a total of 517 articles for consideration (Step 14); the 203 drug citations (Step 26) are a subset of the 517. This search was updated in November 2002, yielding 41 additional articles (step 29 of Table 2). In addition, we ran a somewhat different MEDLINE search in December 2002 starting with the term periodontal disease (instead of periodontitis); this search (line 30 of Table 2) yielded several dozen ostensibly new articles, but because periodontal

disease includes gingivitis, virtually all proved to be irrelevant to our topic, and we added only two new items. The supplemental EMBASE search (Table 3) identified 68 articles, of which 36 were new. We identified three additional studies through references. Finally, external peer reviewers brought three additional studies (from the international literature) to our attention. In all, we examined a total of 599 articles (cited in the Bibliography [Chapter 8]).

Inclusion/Exclusion Criteria

Table 4 specifies the major criteria that we applied to research studies for deciding on their inclusion or exclusion. We applied some in our searches, but opted to cast the net broadly and mainly exclude articles through review of titles and abstracts or the articles themselves. From an earlier preliminary search, we found a considerable number of trials in MEDLINE that appeared to be applicable to our question. For that reason, we searched only for reports of primary research described as trials and those with common characteristics of trials such as blinding and randomization. Although we included review articles that reported on similar types of research (to check references for additional articles), we did not incorporate them directly in this report. We excluded letters, commentaries, editorials, clinical case reports, and practice or treatment guidelines from our searches. Finally, we included only human populations in our search and did not review animal studies.

In our manual review of the articles identified and obtained, one of the senior investigators made an initial decision about inclusion for abstraction and another senior investigator reviewed that decision. In addition, the criteria for inclusion/exclusion were printed on the abstraction form so that abstractors could, if necessary, call for another senior review of eligibility. We determined whether the adjunctive therapy was antimicrobial and chemical at this point; we dropped articles about use of lasers (used presumably to destroy microbes but not a chemical or antimicrobial agent), and those about anti-inflammatory agents (clearly a chemical agent but not antimicrobial). Very late in the process, we also excluded articles reporting on sub-antimicrobial doses of doxycycline, clearly an antibiotic but, at such doses, not intended as an antimicrobial agent.

Many of the studies we identified in our search were not investigations directly relevant to our key question. Many were studying whether a chemical antimicrobial was as effective as SRP in treating periodontitis. However, several studies had a combined (SRP and antimicrobial) treatment group in addition to an SRP-only or SRP-with-placebo treatment group, an antimicrobial-only treatment group, and a no-treatment control group. If the investigators reported comparisons and outcomes of the combined treatment and SRP-only groups for measures we focused on for this report, we included the study. Further exclusions involved studies in which (a) the SRP provided to the group receiving the adjunctive therapy was not the same as that given to the SRP-only group, (b) the periodontitis being treated was not of the chronic (or adult) type but rather a more aggressive variety (e.g., juvenile, early onset, refractory), and (c) the population with the periodontitis had a complicating comorbidity (e.g., HIV/AIDS or diabetes) or a risk factor (e.g., smoking).

We included articles in languages other than English in our *searches* because often the titles and abstracts are translated into English; this allowed us to assess whether we were missing potentially important areas of work. If the full articles were not in English, however, we did not include them in the review. This approach may have caused us to omit some materials in other languages, but our previous systematic reviews have shown that relevant studies done outside of English-speaking nations that would otherwise have met our inclusion criteria would likely have been published in an English-language journal. Thus, we do not think that restricting the full review on this key question to English-language documents introduced any serious bias.

Title, Abstract, and Article Review

To narrow the literature identified through the search to studies with evidence that bore directly on the key question, two senior analysts independently reviewed titles and abstracts obtained in the initial searches. The reviewers were not blinded in any way to authors, journals, or affiliations. After some discussion of the inclusion/exclusion criteria, they retained research studies believed to be focused on the key question.

Of the 517 titles and abstracts independently examined at the first stage, both senior analysts agreed to retain 107 for further review. When they disagreed at this stage, they obtained the full article for review unless it was a foreign article or a review; this step led us to retain an additional 70 citations. Thus, we retained a total of 177 articles to be fully reviewed and possibly abstracted. In addition, we obtained six seemingly relevant review articles. The 36 articles from the EMBASE search and the 43 additional titles and abstracts identified through MEDLINE in November and December 2002 received the same type of review. Of the total 599 articles reviewed, we retained 67 studies to carry through to full article review, abstraction, and inclusion in this report. Finally, as noted, we added three studies that were brought to our attention during the peer review in May 2003.

Data Abstraction

For all retained articles, we obtained hard copies of the full articles. Meanwhile the project team developed a draft data abstraction form and tested it on a small number of articles. We trained abstractors on the initial forms, but the complexities and poor presentation of some of this literature dictated that we revise the abstraction form somewhat to make the process easier. Abstractors were then given an updated training session on the final abstraction form (Appendix B). Ultimately, data items included study identification information, design, descriptions of the sampling and characteristics of the treatment and control groups, description of the adjunctive interventions, reported outcomes and statistics, and other information or comments needed to characterize the study adequately. During this period, the project director, EPC Co-Director, and other EPC staff developed a tentative list of evidence tables, created draft evidence tables with provisional column headings, and established conventions for the order of entry of articles into those tables.

Two senior investigators trained the five main data abstractors, all with master's degrees in public health or another relevant master's degree. Training consisted of a thorough review of the key question, the inclusion/exclusion criteria, and the abstraction form, as well as a walk-through review of up to three articles and independent reviews of two others. We compared the independently abstracted articles to abstracts by the scientific director, noted variations, and provided additional training as needed.

We performed only single abstractions. The EPC's document preparation specialist then entered data from all completed data abstraction forms into the draft evidence tables. For quality control, early in the abstraction process the project director reviewed a small sample of each abstractor's completed data abstraction forms against the full articles; in addition, the EPC's administrator proofread all evidence table entries against the original articles so that needed corrections could be made immediately. Because of the number and size of the evidence tables for this substantial set of articles, we created numerous "text tables" for Chapter 3 that would summarize critical outcome information in a simpler format for users of the report. These tables

also helped us streamline collection of the data needed for the meta-analyses. The EPC administrator also proofread these text tables against both evidence tables and articles, to ensure absolute consistency.

Meta-Analysis

The studies in this evidence report were exclusively clinical trials, albeit many were small and underpowered as individual studies. Investigators often reported the two clinical outcome measures of interest - probing depth (PD) and clinical attachment level (CAL) - in ways that would allow us to express effect sizes in the same way (as millimeters of change). These conditions permitted us to consider performing a series of meta-analyses for studies of at least some of the adjunctive therapies, so that we could quantitatively summarize the work and calculate an overall effect size measure.

We had a total of 70 studies that were candidates for inclusion in one of the meta-analyses. Of those, 29 studies, five with multiple arms, met the criteria that we established for inclusion in the meta-analysis:

- The study had to provide a measure of the treatment effect at 6 months post-baseline, although we allowed for a 3-month window on either side of the 6-month point. Thus, we included studies reporting results from 3 to 9 months (i.e., 12 to 39 weeks). For some articles, the true endpoint of the study fell within this range and they were included; for studies lasting longer than 9 months, the authors had to have reported results from some point within this range; and studies shorter than 3 months were automatically excluded.

- Included studies had to indicate treatment effect by either PD change (i.e., reduction) CAL change (i.e., gain), or both.

- Included studies had to indicate the between-group difference in means (the treatment effect), the standard error or the 95 percent confidence interval of the treatment effect, and the sample size for each study group. Studies could also be included if they provided enough information to allow us to calculate these numbers, such as the within-group differences and their standard deviations or standard errors. Some studies gave the difference between the mean differences of the two study groups (i.e., the effect size of the adjunctive treatment), but often we had to calculate the overall effect size from the mean differences between the baseline to follow-up means for the experimental treatment group and the control group. Similarly, we often had to estimate the standard error of the difference of the mean differences.

- To proceed with a meta-analysis of a given therapy, we required that at least three studies related to one of the clinical outcomes meet the above criteria. The studies of systemic minocycline, systemic metronidazole, and combined metronidazole and amoxicillin failed to meet this requirement.

Comparisons

Regardless of the number of treatment groups and variations in multiple control groups, we narrowed our definition of "treatment" to refer to those groups that received SRP plus one of the following adjunctive therapies that had at least three eligible studies:

- Tetracycline used in systemic therapy and in local applications including fibers, gels, pastes, rinses, solutions, and strips;
- Minocycline used in local applications as a gel, ointment, or microencapsulated;
- Metronidazole, local applications of metronidazole gel; and
- Chlorhexidine used as a local application (e.g., chips, gels, strips, irrigants, and rinses).

Control groups received SRP alone, and no adjunctive drug therapy, other than placebos.

Outcomes

We chose to examine two clinical outcomes that map to our qualitative analyses: PD reduction and CAL gain. For each therapy we analyzed the extent to which the treatment led to a difference in the mean PD reduction or a difference in the mean CAL gain when compared to SRP alone. Our outcome is, therefore, the *difference of a difference* (the difference between the baseline and end-point between the treatment and control groups) or a net between-group difference.

Preparing Study Data for Meta-Analysis

We used the RevMan 4.2 software package to conduct the meta-analysis.[29] RevMan is a software tool designed to manage the entire systematic review process; we used it here exclusively for the meta-analyses. We entered the following information for each study: (a) study ID (author citation) and year; and (b) study design information in the study characteristics table, including methods, participants, interventions, outcomes, and a rating of the allocation concealment or blinding. In the "comparison tables" section, we set up the comparisons (five for PD and four for CAL), listed the outcomes of interest with each comparison, and added the relevant studies to each comparison-outcome node.

Analysis

Analysis was based on the general inverse variance method of estimation available in RevMan. This method calculates a pooled, or overall, effect for each outcome, a test of significance for the treatment effect across all studies (a Z statistic), and a measure of heterogeneity. The heterogeneity statistic is a rough indicator of whether all included studies are indeed comparable (null hypothesis is that treatment effect does not differ among trials). The heterogeneity statistic is also used to calculate an I^2 statistic, which indicates, approximately, the proportion of total variation in the study estimates that can be attributed to heterogeneity rather than sampling error.[29] The method of estimation can handle situations in which effect is specified as fixed or random; we modeled both but have reported only fixed effects for our analysis, as the results were quite similar.

Eight studies did not provide required information on certain aspects of their results. Typically this meant that, to include them, we had to calculate or estimate standard errors ourselves.[30-37] Rather than risk dramatically overestimating the "real" standard error, possibly by as much as 100 percent or more, and unfairly reduce the likelihood of finding statistical significance, we used a set of statistical procedures incorporated in SAS macros[38] to estimate the standard errors of the difference of differences to include in the meta-analysis.

Quality Rating of Articles

To grade the quality of individual studies (articles), as is expected for AHRQ systematic reviews,[39] we developed a quality rating checklist for articles that dealt with internal validity, external validity, and analytic dimensions (Appendix C). The 13 items in the checklist relate mostly to study design elements. We customized items on the checklist to fit the question and literature, but many of the component items were taken directly from, or represent slight modifications of, existing rating scales used by the RTI-UNC EPC, reflecting suggestions from work done by this EPC. [39,40] CONSORT criteria also figured prominently in our thinking because the studies were trials of various kinds.[41]

We pretested draft forms on several articles and eliminated or reworded some items. We gave scores to articles by summing the number of items on the quality rating form checked as "yes" and dividing by 13, the number of items. The EPC Administrator and Project Director independently assigned quality grades to all studies using the final form (Appendix C). We entered both quality scores (essentially two percentages) into the evidence tables. Although not formally validated, our rating scheme adopts the basic strategy of quality grading and provides a relative basis by which we and others can assess the overall strength of the research available to address our key question. We do not employ the quality score as a way of reviewing articles for inclusion in the evidence report or the meta-analysis. This approach is in accord with what recent research has found: no reliable relationship between overall quality rating measures and estimates of treatment effects in trials.[42]

Development of the Evidence Report

Following completion of evidence and/or text tables, we sent them to main authors of the report, together with a general outline of the results chapter of the report. The authors had previously agreed to present mainly qualitative syntheses of the information in the tables, with primary attention to PD and CAL findings and data on spirochetes when available; they would call out information on specific articles only when those studies offered clinically significant findings or insights into the key question. Authors writing from either text or evidence tables were also asked to check data in the tables against articles whenever any table entry was unclear or inconsistent across tables. The authors returned their sections to the project director and EPC Co-Director, who developed an overall synthesis of the results and the discussion chapter. In addition, the project director and other members of the project team developed the research agenda chapter, drawing on the limitations and gaps in the existing literature and on promising leads from the studies reviewed.

We submitted the draft evidence report for external peer review in mid-April 2003 (see Appendix A). Upon receipt of reviews, the EPC staff compiled them into a peer reviewer matrix, discussed many issues with AHRQ and NIDCR staff, and then revised the report as

appropriate, documenting in detail the disposition of all significant clinical, analytic, or policy-relevant changes. We expanded the meta-analyses somewhat to include more studies where we were able to use the SAS macros to produce more estimates of standard errors than we had been able to do before peer review. As noted, we also added several recent articles mentioned by peer reviewers that met our inclusion criteria but had not appeared in any MEDLINE or EMBASE searches. The revised version of the report was submitted to AHRQ and NIDCR for further review before it was put into absolute final form.