Chapter 2: Methodology

Recruitment of Experts

The EPC team identified a core group of 20 technical and community experts to provide input at key points during the project (see Appendix A). These included hepatitis specialists and other experts drawn from academic settings, from relevant professional organizations, and other settings.

The experts from relevant professional organizations were drawn from the American Association for the Study of Liver Diseases, the American College of Physicians-American Society of Internal Medicine, the Infectious Disease Society of America, and the American Academy of Pediatrics. In addition, there was an expert in the assessment of diagnostic technologies and a representative from the Centers for Medicare and Medicaid Services.

The EPC team also identified representatives from a range of other stakeholder organizations to serve as peer reviewers of the draft Evidence Report. The reviewers included an allied health professional, another expert in the assessment of diagnostic technologies, and other clinical specialists drawn from academic and governmental settings (see Appendix A).

The EPC team involved the core experts in defining the key questions (see Identifying the Specific Questions, below) and asked both experts and peer reviewers to review the draft report (see Peer Review Process, below).

Target Population

The targeted clinical population consisted of patients with chronic hepatitis C. The main targeted users of the report are members of the expert panel that is responsible for formulating the consensus statement of the NIH Consensus Development Conference on Management of Hepatitis C in June 2002. This report also should be of interest to clinicians treating HCV-infected patients.

Identifying the Specific Questions

In July 2001, representatives of the EPC team attended the meeting of the Consensus Development Conference Planning Committee that was appointed by the NIH Office of Medical Applications of Research (OMAR). At this meeting the Planning Committee discussed the questions that should be addressed in the Consensus Development Conference and decided on the questions that warranted a systematic review of the literature. The EPC team then formulated these key questions in specific terms that would focus the review process on the most relevant published studies. The EPC team carried out preliminary literature searches and on the basis of those results further refined the key questions. The proposed questions were sent to the core technical experts to rate their relevance and importance. Reviewers commented on the clarity of each question and the availability of evidence to answer it. The EPC team reviewed the experts'

ratings and comments and established the final list of key questions that would be addressed in the Evidence Report.

Key Questions

The EPC team sought to address the following key questions as they pertained to management of chronic hepatitis C.

Role of Initial Liver Biopsy

Q1b: How well do the results of initial liver biopsy predict outcomes of treatment in patients with chronic hepatitis C, taking into consideration patient characteristics such as viral genotype?

Initial biopsy means the biopsy that occurs at initial evaluation before treatment decisions are made. The main outcomes of interest are virologic and histologic measures of disease activity and progression.

Q1e: How well do biochemical blood tests and serologic measures of fibrosis predict the findings of liver biopsy in patients with chronic hepatitis C?

We were interested primarily in biochemical and serologic tests that clinicians could use to estimate the likelihood of fibrosis in patients with chronic hepatitis C. Such information could help guide clinical decisions about the need for an initial liver biopsy.

Treatment Options

Q2a: What is the efficacy and safety of current treatment options for chronic hepatitis C in treatment-naive patients, including peginterferon plus ribavirin, peginterferon alone, standard interferon plus ribavirin, and standard interferon plus amantadine?

Efficacy was assessed in terms of virologic and histologic response to treatment as well as other clinical outcomes including the incidence of cirrhosis, hepatic decompensation, HCC, death, and adverse effects of treatment.

Q2c: What is the efficacy and safety of current interferon-based treatment options (including interferon alone) for chronic hepatitis C in selected subgroups of patients, especially those defined by the following characteristics: age less than or equal to 18 years, race/ethnicity, HCV genotype, presence or absence of cirrhosis, minimal versus decompensated liver disease, concurrent hepatitis B or HIV infection, nonresponse to initial interferon-based therapy, and relapse after initial interferon-based therapy?

Efficacy was assessed in terms of virologic and histologic response to treatment as well as other clinical outcomes including the incidence of cirrhosis, hepatic decompensation, HCC, death, and adverse effects of treatment.

Q2d: What are the long-term clinical outcomes (greater than or equal to 5 years) of current treatment options for chronic hepatitis C?

The main outcomes of interest were the incidence of cirrhosis, hepatic decompensation, HCC, and death. This question included studies of the natural history of chronic hepatitis C because observation is an option.

Screening for Hepatocellular Carcinoma

Q3a: What is the efficacy of using screening tests for hepatocellular carcinoma to improve clinical outcomes in patients with chronic hepatitis C?

The review on this question focused on the following tests: alpha fetoprotein, other serological markers, ultrasonography, computerized tomography, and other imaging studies. The main outcomes of interest were mortality and the rate of resectable versus nonresectable HCC.

Q3b: What are the sensitivity, specificity, and predictive values of tests that could be used to screen for hepatocellular carcinoma (especially resectable carcinoma) in patients with chronic hepatitis C?

The review on this question focused on the same screening tests listed under question 3a.

Causal Pathway

To show how the key questions relate to the overall management of patients with chronic hepatitis C, the EPC team developed a description of a causal pathway (Figure 1). The causal pathway depicts patient characteristics and the types of outcomes that need to be considered in management decisions. The pathway also provides a conceptual framework for identifying gaps in our knowledge about management of chronic hepatitis C.

Literature Search Methods

The literature search consisted of several steps, including identifying sources, formulating a search strategy for each source, and executing and documenting each search. The literature search was conducted through DIALOG, a commercial database vendor, by which each database was searched and the results combined to identify duplicate citations through the DIALOG duplicate checker. This process which delivers a consolidated and nonduplicated list of references, is an economical approach to database searching since the DIALOG system pricing is based on the number of citations downloaded or printed.

Sources

Several literature sources were used to identify all studies potentially relevant to the research questions. Both electronic database searching and manual searching was performed. Eight electronic databases were searched through DIALOG for the period from January 1, 1996 to September 30, 2001. An updated search was completed in March, 2002. The databases

searched are described below. For the key questions that were not addressed at the 1997 Consensus Development conference (i.e., questions 1b, 1e, 3a, and 3b), we also searched MEDLINE® back to 1985.

MEDLINE®

MEDLINE®, or MEDLARS® on-line, is a database of bibliographic citations and author abstracts from approximately 3,900 current biomedical journals published in the United States and 70 foreign countries, dating back to 1966. MEDLINE® was accessed through PubMed, the Internet access to MEDLINE® provided by the National Library of Medicine (NLM).

Biological Abstracts - BIOSIS Previews®

Biosis previews contains citations from *Biological Abstracts*® (BA) and *Biological Abstracts/Reports, Reviews, and Meetings*® (BA/RRM) (formerly *BioResearch Index*®), the major publications of BIOSIS®. Together, these publications constitute the major English-language service providing comprehensive worldwide coverage of research in the biological and biomedical sciences. *Biological Abstracts* includes approximately 350,000 accounts of original research yearly from nearly 6,000 primary journal and monograph titles. *Biological Abstracts/RRM* includes an additional 200,000 citations a year from meeting abstracts, reviews, books, book chapters, notes, letters, selected institutional and government reports, and research communications. U.S. patents are included from 1986 through 1989.

Science Citation Index - SciSearch®

Science citation Index-Sci Search is an international, multidisciplinary index to the literature of science, technology, biomedicine, and related disciplines produced by the Institute for Scientific Information® (ISI®). SciSearch contains all of the records published in the Science Citation Index®, plus additional records from the Current Contents® publications. SciSearch indexes all significant items (articles, review papers, meeting abstracts, letters, editorials, book reviews, correction notices, etc.) from approximately 4,500 major scientific and technical journals. Some 3,800 of these journals are further indexed by the references cited within each article, allowing for citation searching. An additional 700 journals indexed have been drawn from the ISI Current Contents® series of publications.

Manual, Alternative and Natural Therapy™ - MANTIS™

Manual, Alternative and Natural Therapy™ (MANTIS™) is a bibliographic database that provides coverage for health care disciplines not significantly represented in the major biomedical databases. International in coverage, the database contains references from more than 1,000 journals, with preference given to peer-reviewed journals. Approximately 70% of the references have abstracts. Searchable subject headings include Medical Subject Headings (MeSH®), plus a specialized supplemental controlled vocabulary in the areas of alternative medicine.

Allied and Complementary Medicine Database

The Allied and Complementary Medicine Database, formerly the Allied and Alternative Medicine Database, collects abstracts from over 400 biomedical journals as well as articles from other journals that deal with the topic of allied and alternative medicines. Established in 1985, this database is supported by the British Library Healthcare Information Service. After 1997, this database collected information concerning palliative care in addition to continuing to gather information about allied and alternative medicines.

CAB Health

The CAB International resource database, CAB Health, was established in 1973 and is updated quarterly. This database contains citations from both English and foreign language medical writings from over 130 countries. It has gathered over 500,000 citations relating to nutrition, protozoology, medical and veterinary entomology and mycology.

PsycINFO

The American Psychological Association's resource database PsycINFO contains citations and summaries of journal articles, book chapters, books, and technical reports, as well as citations to dissertations, in the field of psychology and psychological aspects of related disciplines, such as medicine, psychiatry, nursing, sociology, education, pharmacology, physiology, linguistics, anthropology, business, and law. Journal coverage, spanning 1987 to the present, includes international material selected from more than 1,300 periodicals written in over 25 languages. Current chapter and book coverage includes worldwide English-language material published from 1987 to the present. Over 55,000 references are added annually through regular updates.

Sociological Abstracts

Sociological Abstracts (SA) covers the world's literature in sociology and related disciplines in the social and behaviorial sciences. Over 1,600 journals and other serial publications are scanned each year to provide coverage of original research, reviews, discussions, monographic publications, panel discussions, and case studies. Conference papers and dissertations are also indexed in the file.

To ensure a comprehensive literature search, the EPC team also examined the reference lists from our database of reference material previously identified through the electronic searching, discussions with experts, and the article review process. In addition, the EPC team reviewed the list of journals that were cited most frequently in the literature searches and nominated additional journals likely to contain relevant articles (see Appendix B). The team reviewed the tables of contents of these journals for all issues published in 2001.

Search Terms and Strategies

The search strategies were designed to maximize sensitivity and were developed in consultation with team members. Preliminary strategies were developed to identify key articles. Using key articles determined to be eligible for review, the team developed and refined search

strategies in an iterative process. A strategy was first developed for PubMed. This strategy was then modified to create separate search strategies for each electronic database (see Appendix C).

Organization and Tracking of Literature Search

The results of both the MEDLINE® and DIALOG searches were downloaded from electronic sources, if possible, or manually entered into a ProCite database. The duplication check in the bibliographic software was used to eliminate articles already retrieved. This ProCite database was used to store citations and track search strategies and sources. The use of this software also allowed for the tracking of the abstract review process.

Abstract Review

As a first step in the review process, two members of the study team independently reviewed the titles identified by the search to exclude those that obviously did not meet our eligibility criteria:

1. written in English
2. includes human data
3. original data
4. information relevant to the management of hepatitis C
5. reports not only basic but also clinical sciences
6. applies to one of our key questions.

Excluded were:

7. meeting abstracts (no full article for review)
8. other incomplete reports (e.g., all data reported in a subsequent publication).

Differences between the two reviewers were adjudicated by other team members. Titles deemed not relevant by both reviewers were excluded from the abstract review process.

All remaining citations were included in the abstract review process. The EPC team developed an abstract review form (Appendix D) to screen the abstracts for relevance. This form was based on forms used in previous EPC reports. Each abstract was circulated to two members of the study team who independently reviewed the abstract and indicated which, if any, of the key questions the article addressed. For articles found not eligible, the reviewers indicated a reason for exclusion. When there was no abstract or when the reviewers could not determine from the abstract whether the article met the eligibility criteria, the team obtained a full copy of the article to review. Disagreements between members of the study team about eligibility were adjudicated at face-to-face meetings.

The EPC team applied the same eligibility criteria at the abstract review phase as listed above.

Qualitative and Quantitative Data Abstraction

The study team developed article review forms that were pilot tested and revised prior to use. The forms included a quality assessment form, a content abstraction form, and supplemental content abstraction forms for the biopsy and screening questions. On the quality assessment form, the reviewers indicated which of the key questions were addressed in the article.

To make sure that all articles met eligibility criteria, the study quality form began with a check of the eligibility criteria (see Abstract Review, above). In addition to the exclusion criteria listed on the abstract review form, the study quality form had other exclusion criteria that were used to focus the search on the studies that would be most valuable in addressing the key questions. These additional exclusion criteria included 1) all data reported in a subsequent publication; 2) some data related to a key question, but the study was not designed to address the question; 3) management of hepatitis C addressed in liver transplant patients only; 4) total number of study subjects less than 30 (as very small studies tend to be less rigorous and were not likely to provide enough valuable data to justify the extra effort needed to extract details from such studies); and 5) outcomes/results not measured according to an appropriate objective standard (i.e., virologic and/or histologic measures for questions 1b, 2a, 2c, and 2d; and histologic or pathologic evidence of HCC for questions 3a and 3b).

In our review of studies on key question 1b (relation of pre-treatment liver histology to outcomes of treatment), we included only randomized controlled trials because they would provide the strongest evidence on whether pre-treatment histologic findings are independent predictors of the efficacy of one treatment strategy compared to another. We were particularly interested in determining whether there is any evidence of an interaction effect between pre-treatment histology and the treatment regimens considered in key questions 2a and 2c. While cohort studies could provide some evidence of the relation between pre-treatment histology and the response to a given treatment regimen, they are susceptible to selection bias in that patients could be excluded from a cohort on the basis of pre-treatment histologic findings. This type of selection bias would make it difficult to determine whether the relative efficacy of different treatment regimens depends on histologic findings. For key question 1b, we also required at least 24 weeks of follow-up

For key question 1e, we included only studies that evaluated biochemical blood tests or serological tests that could serve as measures of liver fibrosis. These studies could include other tests, but we did not include studies that examined only other tests such as hematologic tests or radiologic imaging studies.

For key questions 2a and 2c, we included only randomized controlled trials that had a planned length of follow-up that was at least 24 weeks after the end of treatment. For key question 2d, we included only studies that had at least 5 years of follow-up, including studies of natural history without treatment.

For key question 3a, we looked for studies on patients with chronic hepatitis C that had at least 6 months of follow up for comparing one screening strategy to another or to no screening. For key question 3b, we included only studies that reported data on patients with hepatitis C although these studies could include some patients with only hepatitis B or patients co-infected

with HCV and HBV. We excluded studies that focused solely on hepatitis B because the pathophysiology and natural history of hepatitis C differs from that of hepatitis B.

As shown in Appendix E, the quality assessment form included 23 items about study quality in the following categories: representativeness of study population (five items); bias and confounding (four items); description of therapy/management (four items); outcomes and follow-up (five items); statistical quality and interpretation (four items); and conflict of interest (one item). The items in these categories were derived from study quality forms used in previous EPC projects and were modified to fit a focus on diagnostic and treatment issues in the management of chronic hepatitis C. Because of the divergence of issues covered by our key questions, not all items were required for each of the key questions.

The study team assigned each response level a score of zero (criteria not met), one (criteria partially met), or two (criteria fully met). The score for each category of study quality was the percentage of the total points available in each category for that study and therefore could range from zero to 100 percent. The overall quality score was the average of the first five categorical scores. For consistency with previous EPC reports, we did not include the conflict of interest item in the overall quality score because this was a new item that had not been included in the EPC's assessment of study quality in previous projects.

The content abstraction form included items that described the type of study, geographical location, the definition of study groups, the specific aims, the inclusion and exclusion criteria, screening test characteristics, demographic, social and clinical characteristics of subjects, and outcomes or results related to each of the key questions.

In our review of studies on key question 1b, we looked for the following types of data on: 1) univariate and multivariate analysis of pre-treatment histologic characteristics of patients that were associated with treatment response; and 2) treatment response rates in subgroups defined by pre-treatment liver histology. For studies that reported sustained virologic response rates or sustained histologic response rates in two or more treatment arms for two or more subgroups defined by specific histologic variables, we created a two by two or two by "n" table to record that information for each histologic variable. We used the data in these tables to perform multivariate logistic regression analyses that yielded odds ratios (with 95 percent confidence intervals) for the effect of treatment, effect of pre-treatment histology, and effect size of any potential interaction between treatment regimen and pre-treatment histology.

In our review of studies on key questions 2a and 2c (treatment of chronic hepatitis C), we also looked for data on 1) univariate and multivariate analysis of pre-treatment characteristics of patients that were associated with treatment response; and 2) treatment response rates in subgroups defined by HCV genotype. We focused on the latter because of the reported importance of HCV genotype in predicting response to treatment. For studies that reported sustained virologic response rates or sustained histologic response rates in two or more treatment arms for two or more subgroups defined by HCV genotype, we created a two by two or two by "n" table to record that information for the genotype variable. We used the data in these tables to perform multivariate logistic regression analyses that yielded odds ratios (with 95 percent confidence intervals) for the effect of treatment, effect of HCV genotype, and effect size of any potential interaction between treatment regimen and HCV genotype.

Article Review Process

The team reviewed each potentially eligible article identified by the abstract review process. At least one reviewer in each pair had clinical training, and at least one reviewer had training in epidemiology and research methods. One team member was responsible for completing both the quality assessment and content abstraction forms, and the second reviewed and confirmed the material abstracted. Differences between the two reviewers in either quality or content abstraction were resolved by consensus. Reviewers were not masked to author or journal names because previous work has shown that masking is unlikely to make a significant difference in the results of the data abstraction[15] and would have slowed the review process.

The team developed a Microsoft® Access 2000 (Copyright © 1992-9 Microsoft Corporation) database to collect, maintain, and analyze the quality assessment and content abstraction data. This database was also used to produce the evidence tables.

Evidence Tables

Evidence tables were constructed to present the information obtained on each key question. For each key question, the EPC team created a set of four tables, the first presenting basic information about study aims and eligibility criteria, the second presenting selected characteristics of study participants, the third presenting our assessments of study quality, and the fourth presenting selected results most pertinent to the key question.

Evidence Grades

Five members of the EPC team independently graded the strength of the evidence on each key question. If the team members disagreed about an evidence grade, the final grade given was based on the majority opinion. The grading scheme was derived from the scheme used in previous EPC projects.[16,17] For questions 2a, 2c, 2d, and 3a, the grades included the following:

Grade A (strong): Appropriate data available, including at least one well done randomized controlled trial; study population sufficiently large; adequate controls; data consistent; intervention clearly superior, equivalent or inferior to another strategy;

Grade B (moderate): Appropriate data available; study population sufficiently large; adequate controls; data reasonably consistent; intervention data indicate superiority or equivalence of one intervention compared to another; intervention likely to be superior, equivalent, or inferior to another but insufficient evidence to conclude definitively;

Grade C (weak): Some data available; study population reasonably large; data indicate trend supporting benefit (or equivalence) of one intervention compared to another; insufficient evidence to conclude that intervention is likely to be superior, equivalent or inferior to another;

Grade I (insufficient): Appropriate data not available or insufficient number of patients studied.

For questions 1b, 1e and 3b, the grades were as follows:

Grade A (strong): Appropriate data available, including at least one well done study; study population sufficiently large; adequate reference standard; data consistent; test definitively is or is not useful;

Grade B (moderate): Appropriate data available; study population sufficiently large; adequate reference standard; data reasonably consistent; data indicate test is or is not likely to be useful but insufficient evidence to conclude definitively;

Grade C (weak): Some data available; study population reasonably large; insufficient evidence to conclude that test is or is not likely to be useful;

Grade I (insufficient): Appropriate data not available or insufficient number of patients studied.

Peer Review Process

A copy of the draft report was sent to the core technical experts and the peer reviewers, as listed in Appendix A. Each expert/reviewer was asked to comment on the form and content of specific sections of the report, according to their areas of expertise and interest, and was invited to comment on all other parts as well. The EPC team incorporated the reviewers' comments into the final report.