Chapter 2. Methodology

Technical Expert Advisory Group

We identified technical experts to assist us in formulating the research questions and identifying relevant databases for the literature search. The expert panelists, who are listed in Appendix B, included a neurologist specializing in stroke, a neurosurgeon specializing in severe brain injury, a pediatric neurologist with expertise in treating patients with cerebral palsy, and a physician with an HBOT practice. Throughout the project period, we consulted individual members of the technical expert advisory group (TEAG) on issues that arose in the course of identifying and reviewing the literature.

Scope and Key Questions

The specific questions addressed in this report are:

1. Does HBOT improve mortality and morbidity in patients who have traumatic brain injury or nontraumatic brain injury, such as anoxic ischemic encephalopathy?

2. Does HBOT improve functional outcomes in patients who have cerebral palsy? (Examples of improved functional outcomes are decreased spasticity, improved speech, increased alertness, increased cognitive abilities, and improved visual functioning.)

3. Does HBOT improve mortality and morbidity in patients who have suffered a stroke?

4. What are the adverse effects of using HBOT in these conditions?

To identify the patient groups, interventions, and outcomes that should be included in the review, we read background material from diverse sources including textbooks, government reports, proceedings of scientific meetings, and Web sites. We also conducted focus groups and interviews to improve our understanding of the clinical logic underlying the rationale for the use of HBOT. In the focus groups, we identified outcomes of treatment with HBOT that are important to patients, caregivers, and clinicians and examined whether patients, caregivers, and clinicians who have experience with HBOT value certain outcomes differently from those who have not used HBOT. The methods and results of the focus groups are reported in Appendix C.

The following interventions, populations, outcomes, and study designs were used to formulate the literature search strategy and to assess eligibility of studies.

Intervention

- Hyperbaric Oxygen Therapy: any treatment using 100 percent oxygen supplied to a patient inside a hyperbaric chamber that is pressurized to greater than 1 atm; any frequency, duration, and total number of treatments.

Population

- Patients with brain injury from any cause and in any stage (acute, subacute, or chronic).
- Patients with cerebral palsy of any etiology.
- Patients with thrombotic stroke, excluding patients with transient ischemic attack (TIA), hemorrhage (e.g., subarachnoid hemorrhage), or vasospasm.
- Patients with progressive neurologic diseases (i.e., multiple sclerosis, Parkinson's disease, Alzheimer's disease, and chronic cerebral insufficiency), acute infectious processes (i.e., mucormycosis), radiation sensitization of brain tumors, and reports of treating eye damage or sudden deafness were excluded.
- The use of HBOT for approved indications such as acute carbon monoxide poisoning or acute air embolism was also excluded.

Outcomes

We sought articles reporting any clinical endpoint. In general we excluded studies that reported only intermediate outcomes, such as changes in cerebral metabolism or EEG findings. However, we included studies that reported the effect of HBOT on elevated intracranial pressure, an intermediate outcome that is currently a main determinant of treatment in current clinical practice.

Design

- We included studies of human subjects that reported original data (no reviews of studies).
- We used the algorithm in Figure 1 to classify the design of studies. All of the study designs in the figure were included in the review except for non-comparative studies (e.g., case reports).
- Before-after or time-series studies with no control group were included if (a) five or more cases were reported, and (b) outcome measures were reported for both the pre- and post-HBOT period.

Literature Search Strategy

Electronic Database Literature Search

We searched a broad range of databases to identify published and unpublished studies of the effectiveness and harms of HBOT in patients with brain injury, cerebral palsy, and stroke. Each database initially was searched from its starting date to March 2001. Full details of all the strategies, the databases searched, the inclusive dates searched, the software used to search, and the number of citations found that were used in this review are provided in Appendix D.

The databases we searched were:

- MEDLINE
- PreMEDLINE
- EMBASE

- HealthSTAR (Health Service Technology, Administration and Research)
- CINAHL (Cumulative Index to Nursing & Allied Health)
- Cochrane Database of Systematic Reviews
- Cochrane Controlled Trials Register
- DARE (Database of Abstracts of Reviews of Effectiveness)
- AltHealthWatch
- MANTIS (Manual, Alternative and Natural Therapy)
- Health Technology Assessment Database

If only studies found in the large electronic databases are included, a publication bias may arise in the review. Studies with a positive and statistically significant finding are more likely to be published than those finding no difference between the study groups.[84] Because small studies are more likely to have negative results, this bias has also been called "sample size bias."

Excluding "gray literature" is another potential source of bias. The term "gray literature" refers to reports of studies that are difficult to find, largely because they either are unpublished or are published in sources that are not indexed by the large electronic databases. The Interagency Gray Literature Working Group described gray literature as "foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents."[85] Studies found in the gray literature are not inherently lower quality than those identified through electronic methods, although they are more likely to be small and to have inadequate power to show a difference if one exists.

To avoid publication bias, we asked TEAG members to identify additional databases as potential sources of other material, particularly gray literature, meeting abstracts, and conference proceedings, that may not be indexed in other electronic databases such as MEDLINE. They identified the following sources:

- The Undersea & Hyperbaric Medical Society: a large bibliographic database (30,000 records), http://www.uhms.org/library.htm
- The Database of Randomised Controlled Trials In Hyperbaric Medicine, http://hboevidence.com/
- European Underwater and Baromedical Society, http://www.eubs.org/
- International Congress on Hyperbaric Medicine, http://www.ichm.net/
- National Baromedical Services, Inc.

Each organization was contacted regarding searching their databases. A search of the Undersea & Hyperbaric Medical Society database was conducted by its librarian using our search strategy. A search of the Database of Randomised Controlled Trials in Hyperbaric Medicine was conducted online by the principal investigator. A TEAG member provided the proceedings for 11 of the 12 International Congress on Hyperbaric Medicine conferences (Proceeding number 1 is no longer available). National Baromedical Services, Inc., conducted a search of its database and sent a list of titles at the request of one of the TEAG members. The European Underwater and Baromedical Society did not respond to our requests for access to its database.

Hand Searches

The references of all papers were hand searched. In addition, two reviewers independently conducted hand searches of the references from the *Textbook of Hyperbaric Medicine*.[60] One TEAG member provided articles and meeting abstracts from his personal library. These submitted articles and abstracts were also independently assessed for inclusion by two reviewers.

Update Searches

Update literature searching of the electronic databases MEDLINE, PreMEDLINE, EMBASE, CINAHL, the Cochrane Library, and the Health Technology Assessment Database was completed on February 26, 2002, using the same search strategy as used for the initial searches. The results of these searches are summarized in Appendix D. In May 2003, we added eight additional publications brought to our attention by a peer reviewer. Finally, a supplemental search of MEDLINE, PreMEDLINE, EMBASE, and CINAHL was conducted in July 2003.

Management of References

As such a wide range of databases was searched, some duplication of references resulted. To manage duplicate citations, the titles and abstracts of the bibliographic records were downloaded and imported into Reference Manager, Version 9 (ISI ReSearch Soft, USA), a reference management software program. Due in part to the relatively high proportion of meeting abstracts, some studies present data duplicated in another publication. Where this is clearly the case, only one set of data is presented in the Evidence Tables, and the duplicate publications are noted. Abstracts reporting the same data as found in a full paper were not included. Where multiple publications presented different data from a single study, all were included.

Assessment of Papers for Eligibility

Two reviewers (MM and SC) independently assessed each title and abstract located through the literature searches for relevance to the review, based on the intervention, population, outcome, and study design criteria listed above. Due to time and budget constraints, only studies originally published in the English language were considered for review. This decision was made by the funding agency, AHRQ.

We retrieved the full-text article, report, or meeting abstract of all citations that met the eligibility criteria. Independently, two reviewers reapplied the eligibility criteria to these materials. Disagreements were resolved through consensus.

Data Extraction

Extraction of data from studies was performed by one reviewer (MM for head injury and cerebral palsy, and SC for stroke) and checked by a second reviewer (SC and MH for head injury and cerebral palsy, and MM for stroke). Disagreements were resolved through consensus. Data extracted include first author, year, study population, HBOT protocol, other interventions, study

design, number of patients, outcomes measured, baseline and followup details, results, adverse effects reported, and general comments of the reviewers.

Assessment of Study Validity

All trials were assessed using a list of items indicating components of internal validity in a standardized fashion, based on validity checklists developed at the National Health Service Centre for Reviews and Dissemination and by the US Preventive Services Task Force (Appendix D).[86, 87] Internal validity indicates the level of confidence we have in the accuracy (validity) and reliability (or reproducibility) of the results of the study. The internal validity of a study is assessed based on criteria set for a specific study design. In this way, an observational study would not be judged by criteria for randomized controlled trials (RCTs), but rather by criteria that apply to—and can be met by—a good-quality observational study.

For RCTs and nonrandomized controlled trials, the items assessed for internal validity were randomization/allocation concealment (e.g., randomization and concealment procedures, stratification), baseline comparability of groups, timing of baseline measures, intervention, outcome measures, timing of followup measurements (long enough to assess effects), loss to followup, handling of dropouts or missing data, masking, statistical analysis (if any), and general reviewer comments. The rationale for selecting these criteria is as follows:

- **Methods used to ensure comparable groups at baseline.** Some methods of allocating subjects to treatment and control groups are more likely to prevent bias and to result in groups that are comparable at baseline. Randomization, the best method to allocate patients to groups, is most effective if it is *concealed*. (The importance of allocation concealment is discussed in detail in the Results section in reference to controlled trials of HBOT for traumatic brain injury.)

- **Baseline comparability of groups.** The purpose of randomization (or another allocation method) is to distribute prognostic characteristics equally in the treatment and control groups. Effective randomization distributes known as well as unknown prognostic factors in an unbiased manner. We judged studies on how thoroughly they reported baseline characteristics known to affect prognosis and on whether there were baseline differences between the groups. In a small, well-conducted trial, groups may differ in important baseline prognostic factors because too few patients were randomized. In a large trial, even small differences in baseline characteristics raise concern that randomization failed to distribute unknown prognostic factors equally among the groups. When the method used to conceal allocation is inadequate or is not described, such differences may suggest that randomization was subverted or carried out incorrectly.

- **Use of validated outcome measures.** The use of validated, reliable outcome measures prevents bias on the part of persons who assess outcomes. The use of measures that have not been shown to be valid and reliable reduces confidence that the findings are accurate.

- **Masking of outcome assessment.** The investigators who judge whether the patients have improved should not be aware of which patients received the treatment. This

masking or blinding of outcome assessment is important, because strong beliefs about the benefits of the treatment can influence an observer's assessment of a patient's condition.

- **Maintenance of comparable groups.** Exclusion of subjects after randomization, high rates of loss to followup, and failure to include all randomized patients in the analysis of study results can compromise the quality of a study. Including only those patients who completed the study can give an incomplete picture of the effects of the treatment. For example, if 100 patients are treated and 10 respond, 30 do not respond, and 60 quit the study before their response can be measured – and of these, 20 suffer an adverse event and have to quit treatment – the overall response rate to therapy might be as low as 10 percent. If only those patients who finished the study were included in the statistical analysis, it would appear that the response rate is 25 percent (10/40), when in fact it might have been much lower.

For observational studies, items assessed for internal validity were the establishment of a stable baseline (for before-after and time series studies) or the baseline similarity of the compared groups (if a comparison group was included); discussion of or control for potential confounders; exposure measurement (were all subjects given the same HBOT treatment?); other interventions, the use of valid outcome measures; and the timing of followup measurements.

- **Establishment of a stable baseline.** A before-after treatment study or a time series study relies on the premise that the results after treatment are better than could be expected with standard medical care and the passing of time. For a reader to accept this premise, the study must describe thoroughly the baseline condition of the patients, other aspects of care management, the degree of social support, and any other factor that might predict the outcome. The baseline condition of the patients must be established to be stable; otherwise, changes seen cannot be distinguished from an evolving clinical picture. Omission of even one characteristic that could have accounted for the results raises doubt about whether it was really the treatment that is responsible. The baseline assessments should be timed in a manner that is appropriate to the study's circumstances. For example, the baseline assessments would need to be more frequent in a study of patients being treated in an intensive care unit for acute trauma than in a study of patients who have motor, language, and cognitive deficits many years after trauma.

- **Discussion of or control for potential confounding factors.** In conducting an observational study, the investigators should plan to measure factors other than the use of HBOT that could explain the observed results. Such factors include baseline prognostic characteristics, the natural course of the disease, and the use of other interventions.

  Because they do not use randomization to distribute prognostic factors equally among treated and untreated groups of patients, observational studies usually compare groups that have important baseline differences. For this reason, it is important that baseline characteristics be assessed and reported in detail. When baseline differences are apparent, failure to use appropriate methods to control for bias reduces the internal validity of a study.

  In addition to prognostic factors, differences in the intensity and quality of care can also influence the results of observational studies. In observational studies, treatment

regimens are not determined experimentally but rather by the clinician and patient involved, and they may vary widely between and within groups. Because practice styles vary in many ways, not just in the use of HBOT, other interventions may be used differently and they may have their own impact on outcomes. The interventions used in both groups must be described thoroughly. If differences in management styles and the quality of care are not described, or if they are great, it may be impossible to determine the extent to which the observed results are due to HBOT or to other aspects of care.

- **Use of valid outcome measures and masking of outcome assessment.** The use of validated, reliable outcome measures, rather than the investigator's global subjective judgment, is even more important in observational studies than it is in a RCT. In before-after studies and in many other types of observational studies, the patients, their caregivers, and the investigators are *always* aware of treatment status. Although difficult, it is possible to obtain an independent assessment of results by having unbiased observers who did not participate in administering HBOT rate videotaped examinations made before and after treatment.

Based on these criteria, each study was assigned an overall rating (good, fair or poor) according to the US Preventive Services Task Force methods.[87] The definitions of the three rating categories for these types of studies are as follows.

Good: Comparable groups assembled initially (adequate randomization and concealment, and potential confounders distributed equally among groups) and maintained throughout the study; followup at least 80 percent; reliable and valid measurement instruments applied equally to the groups; outcome assessment masked; interventions defined clearly; all important outcomes considered; appropriate attention to confounders in analysis; for RCTs, intention-to-treat analysis.

Fair: Generally comparable groups assembled initially (inadequate or unstated randomization and concealment methods) but some question remains whether some (although not major) differences occurred with followup; measurement instruments acceptable (although not the best) and generally applied equally; outcome assessment masked; some, but not all important outcomes considered; appropriate attention to some, but not all potential confounders; for RCTs, intention-to-treat analysis.

Poor: Groups assembled initially not close to being comparable or not maintained throughout the study; measurement instruments unreliable or invalid or not applied equally among groups; outcome assessment not masked; key confounders given little or no attention; for RCTs, no intention-to-treat analysis.

The discussion of results and conclusions in this report is based on good- and fair-quality studies. Flaws that have bearing on the interpretation of studies included in this review are discussed in the text and can be examined in Evidence Tables 8 and 9. Results of good-quality studies have a high likelihood of being both valid and reliable. Fair-quality studies have important but not fatal flaws in their design or conduct. The category of fair is broad, with some studies that are probably valid and others that are unlikely to be valid, depending on the specific

flaws found and their severity. The inadequacies found in poor-quality studies make the results unreliable.

External validity refers to the applicability of the results of the study to clinical practice. Although criteria for assessing external validity in systematic reviews are not well-defined, a few criteria can be identified. First, the investigators should describe the criteria used to identify eligible subjects for the study. Second, they should report the numbers of patients who were considered for inclusion in the study, the number that met the eligibility criteria, and the number that actually entered the study. Third, they should report the age range, the severity of disease or disability, the prevalence of comorbid conditions, and other sample characteristics that would enable a clinician to assess the applicability of the results to the patient population for which the intervention is intended.

Quality of the Body of Evidence

We assessed whether the overall strength, quality, and consistency of the body of evidence for each key question. This assessment was based on the internal validity and external validity of the individual studies and the coherence of all the pertinent studies taken as a whole. We also assessed whether the body of evidence was sufficient to provide a clear answer to the key question. In this context, the term "insufficient evidence" refers to the fact that important gaps in the available information remain; this term should be taken to mean that the evidence neither proves nor disproves that HBOT is effective.

Synthesis of Results

Results of data extraction and assessment of study validity are presented in structured tables (Evidence Tables 1-9) and also as a narrative description. We considered the quality of the studies and heterogeneity across studies in study design, patient population, interventions, and outcomes to determine whether meta-analysis could be meaningfully performed. If meta-analysis could not be performed, we summarized the data qualitatively. Assessments of individual criteria for each included study are presented in Evidence Tables 8 and 9, along with the summary measure assigned.

Peer Review

The draft document was sent out for peer review to national experts (see Appendix B.) Their comments were reviewed and, where possible, incorporated into the final document. The final document has not undergone a second review by these reviewers.