Biology of Blood and Marrow Transplantation 11:819-822 (2005)

1083-8791/05/1111-0001$30.00/0
doi:10.1016/j.bbmt.2005.06.005

EDITORIAL

The Evolution of the Evidence-Based Review: Evaluating the Science Enhances the Art of Medicine—Statement of the Steering Committee for Evidence-Based Reviews of the American Society for Blood and Marrow Transplantation

Roy Jones,[1] Yago Nieto,[2] J. Douglas Rizzo,[3] Donna Wall,[4] John R. Wingard,[5] Philip McCarthy, Jr,[6] Theresa Hahn[6]

[1]M.D. Anderson Cancer Center, Houston, Texas; [2]Clinica Universitaria de Navarra, Pamplona, Spain; [3]Medical College of Wisconsin, Milwaukee, Wisconsin; [4]Texas Transplant Institute, San Antonio, Texas; [5]University of Florida Health Science Center, Gainesville, Florida; [6]Roswell Park Cancer Institute, Buffalo, New York

Correspondence and reprint requests: Theresa Hahn, PhD, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263 (e-mail: Theresa.hahn@roswellpark.org).

Received June 27, 2005; accepted June 28, 2005

KEY WORDS
Hematopoietic stem cell transplantation • Evidence-based medicine • Methodology

In little more than a decade, evidence-based medicine has evolved from a theoretical—and often controversial—concept into a widely practiced methodology that can enhance the critical processes of medical decision making. Fears that evidence-based medicine would elevate the science at the expense of the art of medicine and handcuff physicians to rigid "one size fits all" practice guidelines have diminished. A thorough, systematic evidence-based review is a powerful tool to assist physicians and patients who otherwise must make choices on the basis of conventional wisdom, hearsay, and piecemeal empirical data.

We have learned much since 1999, when the American Society for Blood and Marrow Transplantation (ASBMT) launched its initiative to conduct evidence-based reviews of blood and marrow transplantation in the treatment of selected diseases. Often, the process has been as valuable for what it cannot tell us as for what it can. A systematic review may reveal a preponderance of conflicting studies or conclude that there simply is not enough empirical evidence to support one recommendation over another. By taking a hard look at the quality and quantity of the science, evidence-based medicine often highlights the gaps in our science and validates the art that every good physician brings to clinical practice—an art based on a synergistic blend of empirical knowledge, clinical experience, and human intuition.

We also have learned that the art of medicine has a role in the review process itself. In practice, reviewers often must grapple with numerous variables that are difficult to quantify objectively yet affect the quality and strength of the evidence. In response, the standards of evidence-based medicine have evolved beyond grading schema that rank evidence on strictly objective criteria. Although these objective measurements remain important, the Society's first evidence-based reviews for diffuse large-cell B-cell lymphoma (DLCL) [1] and multiple myeloma (MM) [2] taught us not to rely on study design alone to define the best evidence. To do so can undermine optimal patient care, such as when a study is compromised by poor methodology or inappropriate interpretation of results. There were some instances, for example, when the independent experts who served on the DLCL and MM panels unanimously agreed to discount some

of the randomized controlled trials (RCTs) because, in the end, they did not sufficiently answer the questions posed.

CHANGING CRITERIA FOR ASBMT REVIEWS

The original methodology called for a periodic assessment of the process of conducting the reviews [3]. In response to an evaluation of the first 2 reviews, the ASBMT Steering Committee for Evidence-Based Reviews in 2004 adopted the following changes in criteria for conducting future reviews.

Focus of Reviews

The evidence-based reviews conducted by the ASBMT thus far (DLCL, MM, and acute lymphoblastic leukemia [ALL]) have considered all relevant aspects of transplantation for each disease state within a single review article. For future topics, the Steering Committee will establish the initial focus for each review and develop a list of questions to be addressed. This approach not only shortens the review process, but also, more importantly, allows us to focus on the questions that are most relevant to today's clinicians and scientists.

Inclusion Criteria

In past reviews, inclusion criteria for evidence were largely determined by each expert panel. For future reviews, the Steering Committee has established 4 standard criteria:

1. *Meeting abstracts and data from non–peer-reviewed journals will be excluded.*
2. *Only evidence from studies published in 1990 or later will be included.*
3. *A minimum of 70% of study subjects must be patients with the disease under review, or study results must be stratified by the disease to be included.*
4. *Studies with fewer than 25 patients will be excluded, unless they will affect treatment recommendations (eg, where no large studies exist or where they are flawed by problems in design, methodology, or reporting of results).*

Methodology

The methodology for the ASBMT-sponsored reviews was established in 1999 according to well-accepted standards for evidence-based medicine. In April 2001, the US Agency for Healthcare Research and Quality sponsored a study of the methods used for systematic reviews. The agency published a critical evaluation of the established schemas for grading the quality and strength of the evidence and reviewed the 20 it determined to be of the highest quality. After reviewing the systems recommended by the Agency for Healthcare Research and Quality, the Steering Committee selected a grading schema for future reviews based on guidelines developed by the Scottish Intercollegiate Guidelines Network (SIGN) of the Royal College of Physicians of Edinburgh [4].

The SIGN criteria most closely resemble the original grading criteria for the ASBMT reviews yet also address a deficiency of our prior grading schema: the lack of an assessment of quality for individual studies within each category of study design. To evaluate methodologic quality, SIGN uses standardized checklists of criteria to rate studies as follows:

++ All or most criteria from the checklist are fulfilled or, when not fulfilled or adequately described, are judged to be highly unlikely to alter the study's conclusions.

\+ Some of the criteria from the checklist are fulfilled; where not fulfilled or adequately described, they are unlikely to alter conclusions.

− Few or no criteria are fulfilled, and those not fulfilled or adequately described are very likely or likely to alter conclusions.

More information about SIGN, including the criteria checklists for rating methodologic quality, can be found at http://www.sign.ac.uk/guidelines/fulltext/50/.

Validating the System

An unanswered question about SIGN was whether the methodology for quality assessment was sufficiently objective to facilitate consistent, unbiased application of the grading criteria by multiple raters. Before formally adopting the new system for the ASBMT reviews, the Steering Committee undertook an interrater reliability study to validate the SIGN methodology.

Of the approximately 180 scientific articles initially selected for inclusion in the ALL review, approximately 10% (n = 18) were randomly chosen by a third party. Four members of the ASBMT Steering Committee acted as raters of these studies and were asked to rank each article according to the SIGN checklist and rating system. The consensus rating among the 4 raters was significantly consistent: 44% of the time, the raters had perfect agreement; 33% of the time, 1 rater differed; 17% of the time, the raters were split evenly; and only 6% of the time, there was no consensus between raters ($P < .0001$; Pearson exact χ^2). The κ statistic can be used only to compare interrater agreement between 2 raters; therefore, the Pearson exact χ^2 was used to compare the consistency of grading among the 4 raters.

HOW SIGN ENHANCES THE ASBMT REVIEW PROCESS

The SIGN criteria are comparable to those that governed the ASBMT's first 2 reviews, and their ap-

plication, in the Committee's opinion, will not change or invalidate the consensus recommendations reached by the expert panels that conducted the DLCL and MM evidence-based reviews. At the same time, the new schema advances the process by:

- Giving due weight to methodologic variables within a study's design that can significantly affect the quality of evidence or affect the study's applicability to a given patient population.
- Presenting a more thorough overall picture of the entire body of available evidence and avoiding the pitfall of overreliance on the conclusions of single studies or types of studies.

The SIGN system also preserves the strengths of our earlier review criteria. The hierarchy of study types recommended by the Agency for Healthcare Research and Quality and widely accepted by experts remains the first step in grading the quality of design and strength of the evidence.

LEVELS OF EVIDENCE

1++ High-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias.
1+ Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias.
1− Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias.
2++ High-quality systematic reviews of case-control or cohort studies; high-quality case-control or cohort studies with a very low risk of confounding, bias, or chance and a high probability that the relationship is causal.
2+ Well-conducted case-control or cohort studies with a low risk of confounding, bias, or chance and a moderate probability that the relationship is causal.
2− Case-control or cohort studies with a high risk of confounding, bias, or chance and a significant risk that the relationship is not causal.
3 Nonanalytic studies, eg, case reports or case series.
4 Expert opinion.

SYNTHESIS OF THE EVIDENCE

After the evidence has been assembled and rated according to the hierarchy of the study design and the quality of the methodology, relevant data are abstracted from studies that meet the inclusion criteria for the review. Relevant data from the individual studies are summarized in text format, and tables summarizing information about study quality (eg, sample size and duration of follow-up) also are created to facilitate expert assessment of the overall direction and weight of the evidence.

CONSIDERED JUDGMENT

Once the evidence has been synthesized according to rigorous objective standards, the art of medicine comes into play as members of the expert panels begin the process of making and grading the strength of their recommendations. Here, the experts' individual knowledge and clinical experience are drawn upon as they consider the evidence—or lack of evidence—to answer the questions posed by the review. To reduce the risk of introducing personal bias into the review, this step relies on the consensus of many individuals who are experts in myriad aspects of the disease under investigation. The expert panel convened to conduct the evidence-based review of ALL, for example, comprises nationally recognized authorities in both pediatric and adult ALL, including those who specialize in transplantation and those whose expertise is in other treatment modalities for the disease.

GRADES OF RECOMMENDATION

The treatment recommendations of the expert panel also are subject to a standardized grading system.

A At least 1 meta-analysis, systematic review, or RCT rated as 1++ and directly applicable to the target population or a systematic review of RCTs or a body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results.
B A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results or extrapolated evidence from studies rated as 1++ or 1+.
C A body of evidence including studies rated as 2+, directly applicable to the target population, and demonstrating overall consistency of results or extrapolated evidence from studies rated as 2++.
D Evidence level 3 or 4 or extrapolated evidence from studies rated as 2+.

IMPLICATIONS FOR PAST AND FUTURE REVIEWS

The SIGN system was adopted for the Society's evidence-based reviews of ALL in children and adults, which respectively appear in this issue and the January 2006 issue of this journal, and will be the standard for subsequent ASBMT reviews. The new system also will be applied to the completed reviews on DLCL and

MM when they are updated as new evidence becomes available.

It is our belief that the new system will enhance the Society's evidence-based review process by addressing the following:

- The many areas of medical science in which randomized trials may not be practical or ethical.
- Concerns that the controlled, randomized trial, although widely accepted as the most robust study design with the least risk of bias to answer questions of effectiveness, may not always be the best evidence to answer other questions or that it may have methodologic flaws that undermine its strength.
- Interpretations of studies that overgeneralize results, thus contributing to misleading expectations about efficacy.
- The limitations of guidelines that grade the strength but not the importance of the evidence; this may result in user confusion and discourage consideration of some low-grade yet significant recommendations [4].

With this modification in the system for conducting the reviews, the Society signals its commitment to the highest current standards of evidence-based medicine. The more thorough and unbiased the review, the more it meets our objective to support patients and providers in the complicated process of choosing treatment options.

At its best, evidence-based medicine advances its field of inquiry and points toward research that will lead to better diagnostic and treatment options. Panelists for the ASBMT evidence- based reviews identify areas where there is insufficient evidence to support treatment recommendations and prioritize the questions that, in their expert opinion, are most important to answer through future research.

As the science of evidence-based medicine evolves, the ASBMT review process also will evolve to keep pace with advances. The ultimate goal, however, remains unchanged: to give every patient access to the treatment option that offers the best chance for survival and a high quality of life.

REFERENCES

1. Hahn T, Wolff SN, Czuczman M, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of diffuse large cell B-cell non-Hodgkin's lymphoma: an evidence-based review. *Biol Blood Marrow Transplant.* 2001;7: 308-331.
2. Hahn T, Wingard JR, Anderson KC, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of multiple myeloma: an evidence-based review. *Biol Blood Marrow Transplant.* 2003;9:4-37.
3. Jones R, Horowitz M, Wall D, et al. Policy statement regarding the methodology of evidence-based reviews in evaluating the role of blood and marrow transplantation in the therapy of selected diseases. *Biol Blood Marrow Transplant.* 2000;6:524-525.
4. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ.* 2001;323:334-336.