SECTION 1: EVIDENCE-BASED QUALITY IMPROVEMENT, PRINCIPLES, AND PERSPECTIVES

The Principles and Practice of Evidence-Based Neonatology

Roger F. Soll, MD, and Lisa Andruscavage, DO

ABSTRACT. Significant strides toward improving the outcome of newborn infants have been observed during the modern era of neonatal-perinatal medicine. One of the challenges that neonatologists and pediatricians face is deciding when to change current practices. New literature is abundant, but it must be critically reviewed and evaluated before changes in practice are made. The evidence-based medicine process can be used to help in changing practices or adopting new practices. It is based on five steps: 1) forming answerable questions; 2) searching for the best evidence; 3) critically appraising the evidence; 4) applying the evidence in practice; and 5) evaluating one’s performance.

This article reviews the five steps of the evidence-based medicine process. The various levels of evidence, including randomized, controlled trials and systematic reviews, are defined and discussed. An example of a critically appraised topic, a practical tool that can be used as an aid in addressing the first three questions of the evidence-based medicine process, is included. Once the evidence has been evaluated, the decision of whether or not to implement a change in individual practice or in institutional guidelines must be made. These decisions are difficult and require a variety of individual and societal factors to be taken into account.

Examples of evidence-based medicine in neonatal and perinatal medicine include the use of antenatal corticosteroids for promotion of lung maturity, the use of surfactant replacement therapy for the treatment and prevention of respiratory distress syndrome, and prophylactic indomethacin for the prevention of intraventricular hemorrhage. Review of these examples provides insight into the strengths and weaknesses of evidence-based medicine. Pediatrics 1999;103:215–224; evidence-based medicine, critically appraised topic, randomized, controlled trials; meta-analysis, systematic review; quality improvement, antenatal steroids, surfactant therapy, indomethacin.

Significant strides toward improving the outcome of newborn infants have been observed during the modern era of neonatal-perinatal medicine. Consistent improvement in neonatal mortality has been observed throughout the past three decades, in part because of advances in neonatal intensive care. These advances include the regionalization of neonatal intensive care, the introduction of assisted ventilation, the introduction of antenatal steroids, and the introduction of surfactant therapy. However, the past several decades have also observed their share of therapeutic misadventure: the epidemic of retinopathy attributable to the indiscriminate use of supplemental oxygen, gray baby syndrome attributable to chloramphenicol, and increased kernicterus attributable to sulfonamide use. Most recently, the introduction of drugs that used benzyl alcohol as a vehicle led to disastrous complications in newborns. The question that faces the neonatal community in the 21st century is, “How will we continue to improve the outcome of newborns, and yet avoid repeating these medical disasters?” Proponents of evidence-based medical practice believe that the application of the principles of evidence-based medicine will lead to the best care of individual patients and the best health care policies. The following article will explore what evidence-based medicine is, how one can practice evidence-based medicine and give practical examples of the use of evidence-based medicine in neonatal-perinatal medicine.

WHAT IS EVIDENCE-BASED MEDICINE?

Evidence-based medicine has been defined as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.” It is hard to disagree with those goals. Those who find fault with the application of evidence-based medicine offer two somewhat contradictory criticisms. Critics state that evidence-based medicine is just “business as usual” and that the principles of evidence-based medicine already form the basis of their practice. Clearly, this is not the case. A review of the Vermont Oxford Network database of infants with birth weights of 401 to 1500 g demonstrates large variation regarding a variety of common practices. Practices such as the use of nasal continuous positive airway pressure (nasal CPAP) and high-frequency ventilation (HFV) vary greatly between centers. In very low birth weight infants, the median utilization rate of nasal CPAP is 46%, but the interquartile range is broad (25%–65%). Similar variation in practice is observed with more recent technological innovations such as high frequency ventilation. The median utilization rate of HFV is 17%, but the interquartile range is 4% to 24%. If these and other practices are based on the “best evidence,” why is our practice so varied?
Other critics argue that evidence-based medicine is impossible to introduce into routine practice. Although it may be difficult to teach “old dogs” new tricks, Sackett and colleagues offer a basic primer in how to practice and teach evidence-based medicine.6 The focus of this primer is not the ivory tower academician, but the practicing physician. Sackett emphasizes five critical steps in practicing evidence-based medicine: 1) forming answerable questions; 2) searching for the best evidence; 3) critically appraising the evidence; 4) applying the evidence in practice; and 5) evaluating one’s performance.

FORMULATING THE QUESTION

In formulating the question, one must pose answerable, focused questions that are explicit regarding the patient or problem being considered, the intervention being considered, the comparison intervention, and the clinical outcome of interest.9 The question should be clinically relevant, and therefore worth the effort to answer. It can deal with any aspect of care: etiology, diagnosis, prevention, treatment, or prognosis. In this article, the focus will be on evaluating the efficacy of new therapies. Questions must be searchable (i.e., having a reasonable expectation that some research has attempted to answer the question) and clinically relevant. It is simple to pose the unanswerable question, but much harder to train an individual to ask the critical question that could lead to changing practice and improving outcome. In the setting of the teaching service, Sackett proposes handing out “educational prescriptions,” in which the team helps formulate the question and assigns a specific team member the task of searching the literature and reporting the validity and applicability of the evidence found to the team.9

WHAT IS THE “EVIDENCE” IN EVIDENCE-BASED MEDICINE?

As practitioners, we must learn to find and evaluate the best evidence on which to base our practice. Frequently, the best evidence available is not the best evidence possible. The quality (strength) of the evidence ranges from systematic reviews of multiple well-designed randomized, controlled trials (RCTs) to the opinion of respected authorities (based on clinical evidence, descriptive studies, or reports of expert committees). Although we may desire the highest order of evidence, in certain situations this is impossible. RCTs are limited in their ability to evaluate the long-term consequences of therapy and are of no use in evaluating issues such as complex processes of care or environmental issues such as exposure to chemical or industrial hazards.10 Even with these limitations, it is critical that clinicians familiarize themselves with the methodologic issues involved in the proper conduct of RCTs.

RCTs AND SYSTEMATIC REVIEWS

RCTs are considered to be the best method for evaluating the effectiveness of an intervention, and therefore are considered the best available evidence. The methodology of RCTs seeks to minimize bias at all points of the study, and thereby gives the most accurate and reproducible estimates of effects. Random allocation of study subjects is essential to minimize bias at the time of study entry (selection bias), and provides the basis for all traditional statistical comparisons used in analysis of the trial results. Bias can also occur after patient allocation. Appropriate trial methodology seeks to reduce bias in all aspects of the study including exposure to the intervention (performance bias), completeness of follow-up (exclusion bias), and measurement of outcomes (detection or assessment bias).11 Methodologic quality does impact on the interpretation of results: in trials in which treatment allocation was inadequately concealed, treatment effects were significantly overestimated.12 A variety of checklists and scales (or scores) have been developed to help assess the methodologic quality of RCTs.13,14 Checklists and scales assess both the methodologic qualities of trial design and trial reporting. All these instruments are useful in reminding the reviewer of the many issues regarding the methodologic quality of trials. However, none of these instruments is universally accepted because of lack of agreement of the meaningfulness of each criterion and the inability to demonstrate a relationship of the score to the magnitude of reported treatment effects.

If RCTs have been accepted as the “gold standard” of clinical research, then the systematic review is clinical research’s “Fort Knox.” Systematic reviews differ greatly from the traditional narrative overview.15 Traditional narrative overviews are most frequently authored by an acknowledged “expert” in the area of interest. The narrative overview has a broad focus and no specific strategy for how material was found or selected for inclusion. The conclusions are qualitative and subject to the reviewer’s bias. Unlike the narrative overview, systematic reviews apply specific research strategies to identify, appraise, and synthesize data from all relevant clinical studies. The systematic review has a focused clinical question. Systematic reviews limit the bias inherent in narrative overviews by conducting a comprehensive search of all potentially relevant articles and using explicit reproducible criteria in the selection of articles for review. The research design and study characteristics of the included trials are appraised, data are synthesized, and results are interpreted. Qualitative systematic reviews summarize the data, but do not perform further statistical analyses. Quantitative systematic reviews, or meta-analyses, are systematic reviews that use statistical methods to combine the results of multiple RCTs. Pooling the results of previous similar RCTs increases the statistical power lacking in a series of smaller trials, enabling clinicians to have greater security in accepting or rejecting treatment differences demonstrated by the trials.16,17 The statistical technique for pooling results is similar to the statistical methods used in analyzing the data from multicenter trials.

Meta-analysis has its critics. Any attempt at pooling results from various studies will not only incorporate the biases of the primary studies, but will add additional bias attributable to study selection and the
inevitable heterogeneity of the selected studies. LeLorier and co-workers compare the findings of 12 large RCTs with the results of meta-analyses conducted earlier on the same topics. Agreement between the meta-analyses and the large clinical trials was only fair (kappa = 0.35; 95% confidence interval [CI] 0.06, 0.64). However, differences in the point estimates between the RCTS and the meta-analysis were statistically significant in only 12% of the comparisons. The discrepancies noted by LeLorier may be attributable to a variety of biases that may be introduced in the meta-analysis. There are several plausible explanations for how a meta-analysis might obtain a positive result that is not confirmed by a subsequent large well-designed RCT. Publication bias, the tendency for investigators to preferentially submit studies with positive results, and the tendency of editors to choose positive studies for publication, skews the medical literature toward favorable reports of treatment. Unless the authors of the meta-analysis have done a scrupulous search of all available resources, these studies will not be located and the meta-analysis stands a good chance of reporting a false-positive finding. This problem is then further compounded by the greater chance that this false-positive finding will itself be published. Meta-analyses can offer false-negative conclusions because of inappropriate study selection. If the studies selected are not groupable (heterogeneous), then the positive effects observed with one specific treatment or in one specific population may be lost. To minimize bias, the authors of the meta-analyses and the readers of the review must demand the same methodologic quality from these analyses as they would from a RCT. It is essential that all meta-analyses include a prospectively designed protocol, a comprehensive and explicit search strategy, strict criteria for inclusion of studies, standard definitions of outcomes, and standard statistical techniques.

**SOURCES OF SYSTEMATIC REVIEWS**

The field of neonatal-perinatal medicine has a wide variety of available resources. Because of the pioneering studies of the National Perinatal Epidemiology Unit (Oxford, United Kingdom), comprehensive systematic reviews of RCTs in neonatal-perinatal medicine have been published. Systematic reviews in obstetrics and neonatal-perinatal medicine have been published in the books Effective Care in Pregnancy and Childbirth and Effective Care in the Newborn Infant. The initial efforts in obstetrics and neonatal-perinatal medicine have now expanded to include a worldwide effort in all fields of medicine known as the Cochrane Collaboration. The Cochrane Collaboration is an international effort to prepare, maintain, and disseminate systematic reviews of the effects of health care. The Neonatal Collaborative Review Group oversees the creation and updating of systematic overviews of RCTs of interventions in the field of neonatal-perinatal medicine. These reviews, along with reviews in other health care fields, prepared by members in 28 other Collaborative Review Groups, are published electronically in the Cochrane Database of Systematic Reviews, along with criticisms and comments of the reviews. The neonatal reviews also appear on a Web site maintained by the National Institute of Child Health and Human Development. These reviews are updated frequently as new evidence from RCTs become available. Other Internet-based resources, including addresses of the Cochrane Collaboration and the National Institutes of Health-sponsored Neonatal Collaborative Review Group are listed in Table 1.

**SEARCHING BIBLIOGRAPHIC DATABASES**

The efforts of the Pregnancy and Childbirth Review Group as well as the Neonatal Collaborative Review Group represent a rich resource to locate the best evidence for a variety of interventions in neonatal-perinatal medicine. However, this does not obviate the need to learn how to search the medical literature oneself. It is important to become familiar with two bibliographic databases, EMBASE and MEDLINE. EMBASE ( Elsevier Science) is the short format of the Excerpta Medica database. EMBASE covers the biomedical literature from 110 countries and is particularly strong in pharmaceutical and toxicologic studies. MEDLINE, the National Library of Medicine’s database, indexes information from over 3900 biomedical journals published in the United States and 70 foreign countries. A variety of MEDLINE search engines are widely available including OVID, PUBMED, etc. Practitioners must learn effective search strategies in MEDLINE, including the use of MeSH headings (the National Library of Medicine’s subject index), publication types (eg, clinical trial, letter, and review), and text words to create the most efficient literature searches. Haynes and coworkers offer simple strategies to locate the best studies of treatment, diagnosis, prognosis or etiology with the greatest precision. Depending on your needs, you may want your search to be more sensitive (including the greatest number of relevant articles, but also including some less relevant articles) or more specific (including mostly relevant articles, but omitting a few relevant articles). Search engines such as PubMed have incorporated these “filters” into specialized search engines (“Clinical Queries using Research Methodology Filters”). PubMed is available free of charge on the Internet.

**APPRAISING THE EVIDENCE**

Once the evidence has been retrieved, one must learn how to appraise the evidence. First, one must
assess the validity of the evidence. The strongest evidence will come from properly conducted large RCTs or methodologically sound systematic reviews of multiple well-designed RCTs. However, the best evidence available may not meet these criteria. Evidence supporting an intervention should be considered in descending order: 1) strong evidence from at least one systematic review of multiple, well-designed RCTs; 2) strong evidence from at least one properly designed RCT of appropriate size; 3) evidence from well-designed trials without randomization, such as single group, pre-post, cohort, time series, or matched case-control studies; 4) evidence from well-designed nonexperimental studies from more than one center or research group; 5) the opinion of respected authorities, clinical evidence, descriptive studies or reports of expert committees (Table 2). Similar criteria have been proposed by the Canadian Task Force on Periodic Health Examination and by the US Preventive Services Task Force. Of note, the recommendations of the US Preventive Services Task Force make no mention of systematic reviews, let alone list systematic reviews as the best source of evidence.

One must also assess the clinical impact of the reported results. Studies frequently report on surrogate outcomes that may or may not be appropriate substitutes for the truly relevant clinical outcomes. Even when clinical outcomes are reported, the presentation may obscure the clinical relevance of the data. Reevaluation of the data may be necessary so that measures that are relevant to clinicians can be evaluated before acting on the evidence in hand. This may include recalculating the statistics so that a relative risk reduction and an absolute risk reduction are obtained. To appreciate the statistical relevance of these data, a 95% CI should be calculated.

THE CRITICALLY APPRAISED TOPIC (CAT)

Sackett and co-workers offer a method by which these data can be summarized and shared using the CAT. The critically appraised topic is a practical, evidence-based evaluation and teaching tool. In the neonatal intensive care unit (NICU), questions often arise about new or old therapies and diagnostic tests. In the academic setting, house staff and medical students rotate through the NICU and questions reappear throughout the years. In our efforts to introduce evidence-based teaching in our NICU, we have expanded the CAT to include a prospective search of known sources of high-quality evidence in neonatal-perinatal medicine and have added a review of methodologic quality. The goal of the expanded CAT is to make the CAT more useful to future users and not just an academic exercise for the individual appraiser. A sample CAT regarding a practical question of NICU management is included in the Appendix.

The CAT summarizes the first three steps in the practice of evidence-based medicine; forming the question, searching the medical literature, and critically appraising the evidence. The first step is to pose a question that will be narrow enough to define a clinically relevant topic, but broad enough so that evidence can be found that will fit the criteria. Each CAT includes the question broken down into components so that both the appraiser and the future user can understand the dimensions of the topic being appraised. The question should be explicit, including the patient or problem being considered, the intervention, the comparison intervention, and the clinically relevant outcome. In the sample CAT, the question is “In a moderately preterm infant (the patient), will prophylactic use of surfactant (the intervention), compared with the selective administration of surfactant to infants with established respiratory distress syndrome (RDS) (the comparison intervention), decrease death or bronchopulmonary dysplasia at 28 days of life (the outcome)?”

The second step is to perform an adequate search for trials related to the question. In the field of neonatal-perinatal medicine, several sources of references and systematic reviews (including meta-analyses) are available. The availability of these resources allows for rapid access to high quality evidence regarding RCTs and allows for quick creation of a minimum data set of appropriate articles to consider in the critical appraisal. After reviewing these standard sources, one should conduct a formal literature search. Previous knowledge of one or two appropriate articles discovered in the review of standard sources can be very helpful. If these articles are not retrieved in a literature search, then the search strategy should be revised. Simple changes in the search strategy may yield the desired literature. In the sample CAT, when the keyword “surfactant” rather than “pulmonary surfactant” was used as the initial subject, many clinical trials of surfactant were omitted from the OVID (MEDLINE) search. The use of the text word “prophylaxis” instead of “prophylactic” also produces a different set of articles. Using the “publication type” limit to retrieve only the clinical trials and meta-analyses will reduce the number of articles in the set substantially. When searching for the evidence, the search strategy should be saved and filed with the CAT to facilitate future reevaluation of the topic. A list of titles and abstracts should be kept with the CAT to give future users a better understanding of the other potentially relevant studies.

After the literature search on the selected topic is performed, the study that provides the best evidence

<table>
<thead>
<tr>
<th>Type</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strong evidence from at least one systematic review of multiple, well-designed randomized, controlled trials</td>
</tr>
<tr>
<td>2</td>
<td>Strong evidence from at least one properly designed randomized, controlled trial of appropriate size</td>
</tr>
<tr>
<td>3</td>
<td>Evidence from well-designed trials without randomization, single group pre-post, cohort, time series, or matched case-control studies</td>
</tr>
<tr>
<td>4</td>
<td>Evidence from well-designed nonexperimental studies from more than one center or research group</td>
</tr>
<tr>
<td>5</td>
<td>Opinions of respected authorities, based on clinical evidence, descriptive studies or reports of expert committees</td>
</tr>
</tbody>
</table>

to answer the question is used to create the CAT. The criteria for choosing a given article for review will vary. Clearly, one would want to choose the highest order evidence (a systematic review or methodologically sound RCT). However, many trials or overviews may be available and circumstances will dictate which trial is chosen. The trial may be the most recent, the largest, the one that best addresses the population of interest, etc. The rationale for choosing the trial should be stated in the CAT.

The CAT template used in our NICU includes a list of questions to help determine if the results of the trial are valid. Ideally, abstracts of the article should be reviewed with these questions in mind before article selection. The questions address study methodology, allowing the appraiser and future users of the CAT to identify the strengths and weaknesses of the study. For clinical trials involving interventions, methodologic issues include the method of treatment allocation, masking of treatment, masking of outcome assessment, and completeness of follow-up.

If the study is methodologically sound, the evidence regarding the clinical impact of the treatment is then evaluated. Clinical trials may use a variety of statistical techniques in reporting their results. Just because a reported difference is “statistically significant” does not necessarily make the finding clinically important. To assess whether the results of a trial are clinically relevant, the appraiser may need to calculate some simple statistics from the study findings.

The control event rate (CER) and the experimental event rate (EER) are the percentages of subjects experiencing the event of interest in the control and experimental groups, respectively. The effect of treatment is given by comparing the event rate in the treated group with that in the control group. Two calculations for expressing the effect of treatment are:

1) The relative risk reduction (RRR) = (CER − EER)/CER.
2) The absolute risk reduction (ARR) = CER − EER.

The RRR indicates the relative, but not absolute, reduction in event rate. The ARK, also known as the event rate difference, indicates the absolute reduction in the event rate. If the overall incidence of an event is very low, the ARR will be very low as well, even if there is a relatively large difference between study groups. For example, if the CER is 0.5% and the EER is 0.3%, the ARR is only 0.2% but the RRR is 40%. From the ARR, the number needed to treat (NNT) can be calculated. The NNT is calculated by dividing 1 by the ARR. The ARR is helpful in demonstrating the importance of the treatment. In the above example, the NNT is 500. This means that 500 infants would need the experimental treatment to prevent the event of interest in 1 additional infant. Although there is a relatively large decrease in the rate of event occurrence, unless the event is devastating (such as death) and the treatment is fairly benign (such as childhood immunizations), the evidence will not be considered important.

Many studies publish $P$ values (a risk) to state statistical significance. Using the 95% CI allows a more intuitive feel for the significance of the data. The 95% CI is based on the sample size, CER, and EER. CIs can be determined for the RRR, ARR, and NNT. If the results are presented as means, a CI can be calculated for the difference between the means. In the sample CAT, the use of prophylactic surfactant resulted in an ARR of death of 1.3%. The 95% CI for the ARR is 0.1% to 2.5%. The ARR could be as low as 0.1% (NNT = 1000) or as high as 2.5% (NNT = 40).

It is up to the clinician to decide, given this information, whether or not to use prophylactic surfactant for patients similar to those studied.

The CAT itself is formatted to facilitate its use as a teaching tool. Creating a spreadsheet on the computer to calculate the required statistics eliminates the fear of statistics as a barrier of completing the CAT. Once a CAT has been created, it can be stored on paper or electronically and reviewed whenever the question resurfaces or new evidence appears.

**APPLYING THE EVIDENCE**

Once the appropriate research articles are in hand (the evidence) and the data have been summarized in ways that are clinically relevant (creating the critically appraised topic), then it is necessary to decide how to integrate this evidence into one’s clinical practice, or in the case of institutions, into practice guidelines. In the case of the individual patient, one must assess whether or not the results of the randomized trial apply to treatment of that particular individual. A variety of characteristics (age, sex, comorbidity, and disease severity) may be substantially different from the characteristics of the patients enrolled in the trial. These differences may make it unclear whether to extrapolate the results of the evidence to an individual patient. For example, one could decide not to administer a new treatment because the individual being considered for treatment would not have met eligibility criteria for the trial. In these situations, Sackett and co-workers ask us to use our common sense.8 They pose the question in reverse, asking “Is my patient so different from those in the trial that the results can not help me make my treatment decision?” In reality, there are few situations in which you would expect that an intervention would produce qualitatively different results in patients who did not strictly fit eligibility criteria. Only in these situations should you consider rejecting the results. Another common mistake in applying the evidence occurs with inappropriate subgroup analysis. If there was some prior reason for expecting differential responses to an intervention among different subgroups of patients, this analysis should be prospectively planned as part of the study. If not, these analyses are nothing more than a “fishing expedition,” it is very likely that differences will emerge, but it is impossible to distinguish between real effects and chance events.20 Unless there is some persuasive biologic reason to believe that the treatment would be totally ineffective or detrimental to your patient (compared with patients enrolled in the study), one should assume a similar direction of effect on your patient’s illness.
Evidence can also be used in the creation of clinical policies or guidelines. In creating clinical policies, one must consider the evidence regarding the impact of the disease, barriers to implementing the clinical policy, safety, acceptability, and cost effectiveness.\textsuperscript{30} Even in situations where the evidence is extensive, formulating a clinical policy is a difficult task. Decisions regarding clinical policies involve a series of compromises, which take both the individual and societal values into account.

**EVALUATING YOUR PRACTICE OF EVIDENCE-BASED MEDICINE**

To gain expertise in the practice of evidence-based medicine, Sackett\textsuperscript{8} recommends that one evaluate one’s own performance and commitment to the principles of evidence-based medicine. In an individual’s practice, this would include evaluating all steps of the process including whether you are posing the right questions, searching the literature for the best available evidence, creating critically appraised topics, and applying these results in your clinical practice. Questions should be readdressed and the literature searched routinely to update ones practice. On the institutional level, one should evaluate whether their practice policies and guidelines have sought the best evidence available and whether they continue to be updated regarding new evidence in the field. A continuous audit of these processes through quality assurance committees or other similar meetings are essential to maintain the highest level of practice.

**EXAMPLES OF EVIDENCE-BASED MEDICINE IN NEONATOLOGY**

There are several examples of evidence-rich areas in neonatal-perinatal medicine where systematic reviews have been conducted and data are available regarding current practice. These include the use of antenatal steroids to promote lung maturity, surfactant replacement therapy for the prevention and treatment of RDS, and prophylactic indomethacin for the prevention of intraventricular hemorrhage. As one examines the evidence in these areas and evaluates current practice, many of the strengths and weaknesses of evidence-based medicine become apparent.

**Antenatal Corticosteroids**

Liggins and co-workers\textsuperscript{31} demonstrated that antenatal betamethasone decreases RDS and neonatal mortality in the offspring of mothers who receive antenatal treatment. Since that initial study, a total of 18 RCTs studying the effect of antenatal corticosteroids on promoting lung maturity have been conducted. More than 3700 infants were enrolled in these studies. Crowley\textsuperscript{32} has conducted a systematic review of these 18 RCTs. The meta-analysis demonstrates a significant decrease in the risk of RDS (typical odds ratio 0.53, 95% CI 0.44, 0.63; typical risk difference \(-0.09, 95\% \text{ CI } \(-0.11, -0.06\)), a decreased risk of intracranial hemorrhage (typical odds ratio 0.48, 95% CI 0.32, 0.72; typical risk difference \(-0.11, 95\% \text{ CI } \(-0.18, -0.05\)), and a decreased risk of neonatal death (typical odds ratio 0.60, 95% 0.48, 0.75; typical risk difference \(-0.04, 95\% \text{ CI } \(-0.06, -0.02\)).

The majority of these studies were conducted before 1990. Sinclair\textsuperscript{33} asks the question “At what point in the history of trials of antenatal corticosteroids for fetal lung maturation was the aggregated evidence sufficient to show that this treatment reduces the incidence of RDS and neonatal death?” To answer this question, Sinclair performed a “cumulative meta-analysis.” Trials were ordered by their date of publication and meta-analyses were performed sequentially. As previously noted, the initial trial of Liggins and co-workers\textsuperscript{31} demonstrated a significant reduction in the risk of RDS and the risk of neonatal death. As each new trial is added to the cumulative meta-analysis, the risk reduction remains statistically significant. The point estimate of the risk reduction changes little with the addition of each new trial; however, the 95% CI narrows giving increased precision to the estimate of effect. One is hard-pressed to justify the need for so many clinical trials. Despite overwhelming evidence from RCTs, utilization of antenatal steroids in very low birth weight infants remained low throughout the 1980s. The lack of acceptance of the data on antenatal steroids was, in part, attributable to inappropriate subgroup analyses. Clinicians were concerned that antenatal corticosteroids were ineffective in twin gestation, male infants, prolonged ruptured membranes and a variety of other clinical situations. The meta-analysis conducted by Crowley\textsuperscript{32} evaluates the effect of antenatal steroids in several of these subgroups, and establishes that antenatal steroids are effective in a broad range of clinical situations, and not affected by issues such as multiple gestation and gender.

The evidence from the systematic review provided the cornerstone of the National Institutes of Health Consensus Statement regarding the use of antenatal steroids.\textsuperscript{34} The Consensus Statement recommends the use of antenatal corticosteroids for women at risk in a broad range of gestational ages with few exceptions. Wright and co-workers\textsuperscript{35} demonstrated improved understanding of the impact of corticosteroid therapy by obstetricians in the National Institute of Child Health and Human Development Network before and after the National Institutes of Health Consensus Development conference. This change in attitude was reflected by increased utilization of steroids within the NICHD Network. In 1994, more than 41% of at risk infants were treated with either a partial or full course of antenatal corticosteroids compared with 15% in 1990.\textsuperscript{36}

**Surfactant Replacement Therapy**

The use of surfactant replacement therapy in the treatment of neonatal RDS is firmly based in basic science and clinical research. RDS is thought to be attributable primarily to a deficiency of pulmonary surfactant, leading to decreased lung compliance and respiratory insufficiency in affected infants. Animal models demonstrated the ability of both natural surfactant extract and synthetic surfactant to improve pressure volume characteristics and respiratory findings associated with surfactant deficiency.\textsuperscript{37} Initial
studies in human neonates demonstrated rapid improvement in oxygenation and ventilator requirements. In the past decade, 33 RCTs involving more than 6000 infants have been conducted. The systematic review of these trials demonstrates that surfactant, whether used prophylactically in the delivery room to prevent RDS or in the treatment of established RDS, leads to a significant decrease in the risk of pneumothorax, and the risk of mortality. These benefits were observed in both the trials of natural surfactant extracts and synthetic surfactants. For example, natural surfactant extract treatment of infants with established RDS will lead to a 57% reduction in the risk of pneumothorax and a 32% reduction in the risk of dying.

Unlike the use of antenatal steroid therapy, surfactant therapy became widely adopted when surfactant preparations became available. Within the first year after the Food and Drug Administration approved surfactants, over 54% of all very low birth weight infants in the Vermont Oxford Network received surfactant treatment. In the case of surfactant replacement therapy, a treatment that was well-supported by both basic science research and clinical research quickly diffused into clinical practice. However, careful examination of the evidence shows that we may not be using surfactant in its most optimal fashion. The initial RCTs compared prophylactic delivery room administration of surfactant or treatment of infants with established RDS to no treatment. Those who favored the use of prophylactic surfactant noted that surfactant distributes better when administered into the fluid-filled lung. Investigators also noted that, in animal models, evidence of bronchiolar necrosis and other lung damage could be observed even after short periods of assisted ventilation. Advocates of treating only infants with established RDS noted that twice as many infants would need to be treated prophylactically, increasing both cost and potential risk. A systematic review of the seven RCTs that have evaluated the use of prophylactic surfactant therapy compared with selective therapy of infants with established RDS suggests that there are important clinical benefits associated with prophylactic surfactant administration. Prophylactic surfactant therapy leads to a decreased risk of pneumothorax (typical relative risk 0.62, 95% CI 0.42, 0.89; typical risk difference −0.02, 95% CI −0.04, −0.01) and a decreased risk of mortality before hospital discharge (typical relative risk 0.75, 95% CI 0.59, 0.96; typical risk difference −0.05, 95% CI −0.09, −0.01). Yet, prophylactic care is not widely adopted. In high-risk infants <1000 g, 20% to 30% do not receive any surfactant therapy. The meta-analysis suggests that for every 100 infants treated prophylactically, there will be 2 fewer pneumothoraces and 5 fewer deaths. Despite the increased exposure to treatment and cost of treatment, these clinical benefits appear great enough to warrant these expenses. An evidence-based practice would strongly consider a policy of prophylactic surfactant administration to high-risk infants, or consider participation in further trials which might test strategies that can identify the infant who would most benefit from prophylactic treatment.

**Prophylactic Indomethacin**

Applying evidence-based medicine to practice becomes more complicated when the value of the outcome is less clear (virtually any time the outcome is something other than mortality) and there are actual or theoretical concerns of competing risk. This is the case regarding the use of prophylactic indomethacin in the prevention of intraventricular hemorrhage. Prophylactic indomethacin has been evaluated both in the prevention of patent ductus arteriosus and in the prevention of intraventricular hemorrhage. Thirty-eight percent of all extremely low birth weight infants experience some degree of intraventricular hemorrhage; 17% of which are of the more severe grades. Animal studies and clinical trials have suggested that indomethacin, a cyclooxygenase inhibitor of prostaglandin synthesis, lowers the risk of intraventricular hemorrhage in very low birth weight infants. Indomethacin has been demonstrated to modulate cerebral blood flow, decrease serum prostaglandin levels, and promote germinal matrix maturation.

In clinical studies, indomethacin promotes pharmacologic closure of the patent ductus arteriosus and reduces the incidence of intraventricular hemorrhage. In follow-up of these infants, investigators have demonstrated a trend toward improved outcome. Fowlie conducted a systematic overview of 14 RCTs of prophylactic indomethacin involving 1000 infants. The meta-analysis suggests a decrease in the risk of patent ductus arteriosus (typical event rate ratio 0.31, 95% CI 0.22, 0.44; typical risk difference −0.22, 95% CI −0.28, −0.16), intraventricular hemorrhage (typical event rate ratio 0.74, 95% CI 0.63, 0.87; typical risk difference −0.08, 95% CI −0.13, −0.04), and severe intraventricular hemorrhage (typical event rate ratio 0.60, 95% CI 0.42, 0.87; typical risk difference −0.04, 95% CI −0.07, −0.01), associated with prophylactic indomethacin administration. In clinically relevant terms, one needs to treat 5 infants with prophylactic indomethacin to prevent one patent ductus arteriosus, 12 infants to prevent one intraventricular hemorrhage, and 26 infants to prevent a severe intraventricular hemorrhage. However, use of prophylactic indomethacin is not widespread because of concern regarding possible side effects of treatment including cerebral ischemia and necrotizing enterocolitis. To answer this question, many investigators have chosen to participate in further large RCTs that will address the issue of neurodevelopmental outcome of infants treated with prophylactic indomethacin. It is impractical to believe that, in all clinical situations, there will be the commitment to the time and expense of conducting extremely large RCTs to follow for neurodevelopmental or other long-term outcomes. Sophisticated models need to be applied to decide whether to accept or reject the evidence. This is an ideal situation in which to apply decision analysis. Decision analysis is useful in situations where there are competing risks, allowing for a probabilistic, quantitative frame-
work to aid in decision-making. 48 Decision analysis requires one to structure the problem, assign probabilities to chance events, assign utility or value to all outcomes, evaluate the utility of each strategy, and perform a sensitivity analysis. Structuring the problem involves creating a decision tree, an attempt to simplify a complex problem while still maintaining its clinical meaning. Even relatively simple decisions become complex quickly. In the case of prophylactic indomethacin, one can define three treatment alternatives; prophylactic indomethacin for all at-risk infants, cranial ultrasound screening for baseline intraventricular hemorrhage and indomethacin administration for at-risk infants without severe intraventricular hemorrhage and indomethacin administration only to infants with symptomatic patent ductus arteriosus. The decision tree incorporates estimates derived from the clinical literature regarding the baseline risks of intraventricular hemorrhage, patent ductus arteriosus, and the theoretical risk associated with indomethacin therapy. It is at this point in the exercise that a thorough search of the literature to access the best evidence is crucial. Data for the estimates should come from the highest order evidence including meta-analyses or large RCTs, as well as estimates from large databases.

The results of the decision analysis help inform clinicians regarding the decision to use prophylactic indomethacin.49 Obviously, if there are no ischemic complications, the decision analysis will support the results of the meta-analysis: prophylactic indomethacin either when given to all high-risk infants or when given to high-risk infants after ultrasound screening for baseline severe hemorrhage will be the preferred treatment strategies. If one postulates serious and as yet unknown complications of indomethacin therapy, for example an ischemic brain syndrome, one can evaluate the treatment threshold given either the clinician’s or the patient’s estimate of the impact of either intraventricular hemorrhage of ischemic brain syndrome on the patient’s quality of life. If we assign a significant quality of life reduction to ischemic complications (equivalent to the reduction of quality of life of a major intraventricular hemorrhage), then the decision flips to symptomatic therapy when the probability of ischemic complications increases above 5%.

Although the probabilities may be objectively derived from high-quality evidence, the perceived value or “utility” will always be subjective and therefore cause different people to act differently based on the same evidence. Decision analysis allows for a variety of parameters to be changed according to the perceived values of clinicians or patients. For example, if the quality of life adjustment factor for ischemic brain syndrome were increased, then the threshold point for the probability for where the decision analysis supports symptomatic therapy would increase as well. Data can be presented in multiway sensitivity analyses, allowing clinicians and patients to examine interactions of combinations of any of the variables (quality of life assessments and incidence of complications). Only with these sophisticated models can one begin to use the available evidence in clinically useful ways.

CONCLUSIONS

The field of neonatal-perinatal medicine has developed significant resources to help support the practice of evidence-based medicine. To begin to use these resources, we must prepare ourselves to become practitioners of evidence-based medicine. This includes mastering the five steps of practicing evidence-based medicine.6 It is essential that we bring our expertise in evidence-based medicine to our patients; however, it will be difficult to practice evidence-based medicine in isolation and unreasonable to think that national guidelines or other institutional guidelines will be available regarding most of our treatment decisions. Although evidence-based practice can be successfully practiced on an individual practitioner and patient level, institutional approaches will help reinforce the practice of evidence-based medicine. One such approach is the “benchmarking” approach used in the Vermont Oxford Network.50 Special interest groups are formed with an interest in specific questions (decreasing chronic lung disease, decreasing nosocomial infection). These groups evaluate the available evidence regarding what practices might lead to improved outcome. Some concepts are strongly based in evidence and others are less well-supported. Interinstitutional collaboration allows for critical appraisal of the evidence as well as the creation of a large database of hospitals, which are not as prone to the expected variations observed in small samples.

In conclusion, clinicians must commit to learning the principles of evidence-based medicine and must help create new evidence through participation in relevant clinical trials.

APPENDIX: Critically Appraised Topic
The Use of Prophylactic Surfactant Versus Surfactant Treatment for Established RDS

The question
The patient or problem being considered: in a moderately premature newborn
The intervention being considered: will prophylactic surfactant administration
The comparison intervention: compared with surfactant treatment of established RDS
The clinical outcomes of interest: decrease death or BPD at 28 days?

Searching for the evidence
2. Cochrane Database of Systematic Reviews*—systematic overview (1997)
3. Effective Care of the Newborn—systematic overview (1992)
4. OVID (MEDLINE): Database 1993–January 1998. (Pulmonary surfactants (exploded)) and (respiratory distress syndrome (exploded)) and (textwords [prophylaxis or prophylactic or prevention]). Limited to (clinical trial or meta analysis) and (English). Twenty-seven articles retrieved.

The study appraised: Prophylactic administration of calf lung surfactant extract is more effective than early treatment of respiratory distress syndrome in neonates of 29 through 32 weeks’ gestation†
Rationale: Largest randomized, controlled trial; the only trial (including systematic reviews) studying only infants ≥30 weeks gestation.

The evidence: Randomized, controlled, unmasked trial. A total of 1398 infants were enrolled and the data for 1248 infants were analyzed. Inclusion criteria included birth to 29 to 32+6 weeks gestation. The surfactant used was Infasurf (calf lung surfactant, 150 mg/dose). The prophylactic group was intubated and received surfactant at delivery. The early treatment group received surfactant only if they met criteria for mild RDS ($P_{aO_2} > 3$ and no other disorder noted on CXR). Subsequent doses of surfactant were given to infants in either group if they met criteria for severe RDS (mean airway pressure > 10 cm H$_2$O + $P_{aO_2} > 0.60$ for $P_{aO_2} < 70$ mm Hg).

Is the evidence from the trial valid?
1) Is the trial randomized? Yes. Consent was obtained before delivery. Randomization was performed via sealed envelope when delivery appeared imminent.
2) Were all the patients who entered the trial accounted for at its conclusion? Yes. One hundred fifty infants were excluded for congenital anomalies (62), infection (45), excess maturity (39), and miscellaneous reasons (4).
3) Were all the patients analyzed in the groups to which they were randomized? No. Twenty-four infants received the wrong treatment and were analyzed by treatment received.
4) Were patients and clinicians masked to which treatment was being received? No.
5) Aside from the experimental treatment, were the groups treated equally? Yes.
6) Were the groups similar at the start of the trial? Yes.
7) Is the hypothesis stated? Yes. To show no difference in the development of moderate RDS whether surfactant is given at delivery or once mild RDS develops.
8) Is sample size computation stated? Yes. One thousand three hundred twenty infants were required for a power of .90 to detect a 7% difference in the development of moderate RDS at the P < .05 level.

Is this evidence about a treatment important? (see text)

<table>
<thead>
<tr>
<th>Event</th>
<th>CER</th>
<th>EER</th>
<th>RRR</th>
<th>ARR</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPD at 28 days</td>
<td>7%</td>
<td>5%</td>
<td>29%</td>
<td>2%</td>
<td>0.6%, 4.6%</td>
</tr>
<tr>
<td>Death</td>
<td>1.8%</td>
<td>0.5%</td>
<td>72%</td>
<td>1.3%</td>
<td>0.1%, 2.5%</td>
</tr>
<tr>
<td>Death or BPD at 28 days</td>
<td>9%</td>
<td>5%</td>
<td>44%</td>
<td>4%</td>
<td>1.2%, 6.8%</td>
</tr>
</tbody>
</table>

Control = Early treatment group
Experimental = Prophylaxis group

Abbreviations CER, control event rate; EER, experimental event rate; RRR, relative risk reduction; ARR, absolute risk reduction; NNT, number of patients that need to be treated to prevent one occurrence (1/ARR).

Clinical bottom lines
1. There was a significant reduction in the number of deaths in the prophylactic group.
2. There was a significant reduction in the number of infants with BPD at 28 days in the prophylactic group.
3. There was a significant reduction in death or BPD at 28 days with prophylactic surfactant administration group.

Comments
1. Although infants were eliminated from analysis for excessive maturity, there is no mention of elimination for insufficient maturity. This may have skewed the results towards prophylaxis.

Appraiser: L. Andruscavage
CAT: Therapy
Date appraised: 6/28/98

REFERENCES


**Updated Information & Services**

including high resolution figures, can be found at:
/content/103/Supplement_E1/215.full.html

**References**

This article cites 34 articles, 8 of which can be accessed free at:
/content/103/Supplement_E1/215.full.html#ref-list-1

**Citations**

This article has been cited by 3 HighWire-hosted articles:
/content/103/Supplement_E1/215.full.html#related-urls

**Subspecialty Collections**

This article, along with others on similar topics, appears in the following collection(s):
- Administration/Practice Management
  /cgi/collection/administration:practice_management_sub
- Quality Improvement
  /cgi/collection/quality_improvement_sub
- Evidence-Based Medicine
  /cgi/collection/evidence-based_medicine_sub
- Fetus/Newborn Infant
  /cgi/collection/fetus:newborn_infant_sub

**Permissions & Licensing**

Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

**Reprints**

Information about ordering reprints can be found online:
/site/misc/reprints.xhtml