

Evidence-Based Practice in Provision of Amplification

Robyn M. Cox*

Abstract

Evidence-based practice (EBP) has been widely embraced in many health-care fields as a way of maintaining currency of knowledge and state-of-the-art treatment recommendations in an age of information abundance and rapid scientific progress. Although the principles of EBP are slowly entering the specialties of communication disorders, they are not well known or extensively employed as yet. In this article, the rationale for EBP is presented and differences between EBP and traditional practice are highlighted. The five-step process of EBP is described: defining the problem, searching for evidence, critically appraising the evidence, formulating a recommendation, and assessing the outcome. Critical appraisal calls for determination of the validity, clinical significance, and applicability of evidence. Each of these topics is explored with emphasis placed on the application of EBP to hearing rehabilitation. Finally some suggestions are offered for researchers and practitioners to facilitate a transition to EBP in amplification provision.

Key Words: Critical appraisal, effect size, evidence-based practice, hearing aid

Abbreviations: CI = confidence interval; CL = compression limiting; EBP = evidence-based practice; ES = effect size; NHST = null hypothesis significance testing; OHC = outer hair cell; RCT = randomized controlled trial; WDRC = wide dynamic range compression

Sumario

La práctica basada en evidencia (EBP) ha sido ampliamente acogida en muchos campos relacionados con la salud, como una forma de mantener actualizado el conocimiento y las recomendaciones de punta en tratamiento, en una época de información abundante y de rápido progreso científico. Aunque los principios de la EBP están introduciéndose apenas lentamente en las especialidades de los trastornos de la comunicación, aún no son conocidos o empleados en forma extensiva. En este artículo, se presenta la justificación para la EBP y se destacan las diferencias de la EBP y la práctica tradicional. Se describe el proceso de cinco pasos de la EBP, con la definición del problema, la búsqueda de la evidencia, la evaluación crítica de dicha evidencia, la formulación de recomendaciones y la evaluación de los resultados. La evaluación crítica exige la determinación de la validez, el significado clínico, y la aplicabilidad de la evidencia. Cada uno de estos tópicos se explora con énfasis en la aplicación de la EBP sobre la rehabilitación auditiva. Finalmente, se presentan algunas sugerencias para investigadores y clínicos para facilitar la transición de la EBP a la prestación de servicios de amplificación.

Palabras Clave: Evaluación crítica, tamaño del efecto, práctica basada en evidencia, auxiliar auditivo

*University of Memphis, Memphis, TN

Robyn M. Cox, Speech and Hearing Center, 807 Jefferson Ave., Memphis, TN 38105; Phone: 901-678-5831; E-mail: robyncox@memphis.edu

Presented at the Jackson Hole Rendezvous: A Conference for Hearing Health Professionals, Jackson Hole, WY, Sept. 8–11, 2004.

Abreviaturas: CI = intervalo de confianza; CL = limitación de la comprensión; EBP = práctica basada en evidencia; ES = tamaño del efecto; NHST = prueba de significación de la hipótesis nula; OHC = células ciliadas externas; RCT = estudio aleatorio controlado; WDRC = compresión de rango dinámico amplio

It is not surprising that there is ongoing concern, among the providers of hearing health care, about the level of effectiveness of fitted hearing aids. Even though hearing aid technology has clearly advanced, the percentage of hearing-impaired people owning hearing aids (about 22%) has not changed since 1991 (Kochkin, 2001). Further, as reviewed by Van Vliet (this issue), overall satisfaction with hearing aids has not improved in the past decade.

Why have better hearing aids not produced a corresponding improvement in satisfaction with amplification? The answer lies, at least partly, in the fact that the scientific basis of hearing aid fitting has fallen far behind the technological development of amplification devices (Medwetsky et al, 1999). This problem has at least two components. First, there is relatively little high-quality research to provide effectiveness guidelines for the fitting process. Second, practitioners are generally not well prepared to critically evaluate the body of research that does exist. As a result of these factors, professionals involved in provision of hearing health care often do not have an accurate appreciation of the value to the hearing-impaired listener of technological developments in amplification or other newly proposed treatments (such as audiological rehabilitation strategies). There is a need to promote the ability of practitioners to recognize the potential of promising new treatments and technologies and to apply them appropriately. At the same time, it is equally important for professionals to be able to identify those proposed hearing care improvements that are less valuable to the user than originally anticipated by their developers.

Hearing health care is not alone in the difficulties it currently faces. Health-care professionals in many fields have encountered problems maintaining an up-to-date knowledge base in the face of rapid technology changes and expanding scientific knowledge

about disorders. The principles of evidence-based practice (EBP) have gained adherents in the past decade as a response to this challenge. EBP offers a new point of view about the responsibilities of researchers, teachers, learners, and practitioners. It recognizes that in an age of information explosion, health-care practitioners must acquire skills that allow them to rely more heavily on their own resources and less on traditional sources of authority to evaluate new developments and treatments, or they cannot maintain currency of knowledge. EBP is an orientation that promotes continuous patient-driven overhaul of treatment protocols to incorporate new knowledge about treatment effectiveness.

Although the principles of EBP have become widely circulated in medicine, other health-care specialties have been slower to embrace this new approach. In hearing rehabilitation, there is a small but growing collection of explicitly evidence-based publications (e.g., Robinson, 1999; Hanratty and Lawlor, 2000; Amlani, 2001; Maki-Torkko et al, 2001; Taylor et al, 2001; Sherbourne et al, 2002). In addition, professional organizations have begun to endorse and promote the adoption of EBP in hearing care and other communication disorder treatments (e.g., ASHA, 2005; Dollaghan, 2004). It is likely that this activity presages a groundswell of enthusiasm for the adoption of EBP in hearing care.

To support this effort, this article informs hearing-care practitioners about the principles and applications of EBP with particular emphasis on its use in hearing aid provision. The goal is to empower practitioners to begin implementing the paradigm shift that is necessary to make EBP a part of their everyday routine. Four questions will be addressed: What is new in evidence-based practice? How do practitioners actually do EBP? How do practitioners evaluate the evidence? What do researchers need to do to facilitate the efforts of practitioners?

WHAT IS NEW IN EBP?

One way to approach this question is to consider current typical practice and then to point out how EBP is different. In typical current practice, we assume that what we learn from professors and lecturers about hearing and psychoacoustical principles gives us an adequate foundation for a professional career. Once in practice, when we need to evaluate new treatments or technologies, we attempt to do this by applying our traditional training and using our own common sense. Additionally, as professional experience builds, we accumulate unsystematic observations about what seemed to work on past patients, and that information is used to generate recommendations for treatments for future patients. Thus, when we decide what amplification to use with a new patient, we tend to use what has seemed to work in the past, rely on information from an authority such as a company representative, or call on traditional expertise from a professor or a textbook. In this mode of practice, there is emphasis on acquiring knowledge from authority figures and combining that knowledge idiosyncratically with our personal clinical experience to produce a course of action with each new patient.

EBP approaches clinical practice from a somewhat different orientation. From this point of view, traditional education is seen as equipping the future practitioner with knowledge about basic hearing mechanisms and psychoacoustics, and facilitating acquisition of good clinical skills. These fundamentals provide important underpinnings that are essential for successful practice. However, when the principles of EBP are used, neither unsystematic clinical experience, nor the opinions of authority figures, nor knowledge of basic hearing disorders is a sufficient basis for determining amplification treatment for a given individual. Instead, the recommendation for each patient should be based on the clinician's insightful appreciation of the goals, needs, and preferences of the individual patient *in combination with* the best available data about the potential treatment. Ideally, the data should reflect effectiveness of the treatment (such as a new hearing aid feature) in real life with patients similar to the current

patient. No matter how logically appealing a new treatment may seem to be, it cannot be assumed to perform as planned until there is specific effectiveness data that verifies this. Unless such data exist, practitioners must acknowledge that there is uncertainty about the value of the treatment to patients (even when data do exist, there is uncertainty about results for a particular patient). Finally, EBP assumes that practitioners can and should learn to evaluate evidence from original research to support their treatment recommendations.

Hearing aid provision frequently produces questions that call for the application of the principles of EBP. For example: Is wide dynamic range compression processing better than linear processing with compression limiting? Do proprietary "fast fit" algorithms generally result in good fitting outcomes? Is a fitting more acceptable if we match maximum output levels to clinical loudness discomfort levels? Do second-order directional microphones give better real-world results than first order? Our answers to such questions (and therefore our treatment recommendations) should not be based on armchair logic or on marketing publicity. Unless there is specifically relevant real-world data to support our assertions, we must acknowledge that we do not know the answers to these kinds of questions.

HOW DO PRACTITIONERS IMPLEMENT EBP?

It is important to realize that evidence-based practice is not a vague overall orientation. It is a well-defined, stepwise activity that is accomplished one patient and one problem at a time. In this process, a treatment problem is identified that can be addressed empirically. The practitioner decides to use the principles of EBP to critically appraise the topic to find the best available recommendation to deal with the problem. There are five distinct steps for application of EBP:

1. Generate a focused, answerable clinical question.
2. Find the best available evidence.
3. Evaluate the validity, importance, and relevance of the evidence.
4. Generate a recommendation by combining evidence with clinical expertise and patient variables.

5. Evaluate the result and seek ways to improve.

Each step is discussed below.

Step 1: Generate the Question

The question comprises several essential elements. It must specify the important aspects of the patient (age, gender, etc.), sufficient detail about the problem (e.g., type/extent of hearing loss), the proposed intervention or treatment (e.g., hearing aid with speech enhancement algorithm), the comparison or alternative treatment, if appropriate (e.g., hearing aid without speech enhancement algorithm), and the evidence that will be regarded as important or convincing (e.g., laboratory measurement of phoneme perception in noise).

Consider the following example: Ethel Tribbet is 75 years old and lives alone on a limited income. She has bilateral moderate presbycusis. She has worn hearing aids for 10 years, mainly for watching television. Her latest pair were programmable analog types with directional microphones. They still work fine. She does not go out much, but her children visit for lunch on Sundays. When they all get together, she has a hard time following the conversation, which is increasingly annoying for everyone. Her daughter has accompanied her to the audiologist’s office. Her family wants to know if the new “digital” hearing aids will help Mrs. Tribbet more than her current fitting. If so, they will get together to buy them for her. If not, they will not buy new hearing aids. They are asking for a recommendation.

After reflection, the practitioner may decide that he/she does not know what patient-oriented evidence is available to answer the family’s questions, and decides to critically appraise this topic. The question generated in Step 1 might be as follows: “Will

an older woman with moderate bilateral presbycusis obtain better speech understanding in noise with digital processing hearing aids than with programmable analog devices with similar performance specifications?” The elements of the question are made explicit in Table 1. It is worth taking time to carefully construct the question because doing so: (1) forces the practitioner to decide exactly what he/she wants to know, (2) focuses efforts in seeking the evidence, which (3) saves time, and (4) facilitates arriving at the best available answer.

Step 2: Find the Best Available Evidence

The principles of EBP require that the evidence on a treatment question should be data based. The data-based evidence may be derived from several sources. Peer-reviewed journal articles are the most obvious place to look for high-quality evidence. This source of evidence has the advantage of having been refined by a prepublication assessment process. There are other excellent sources of evidence, however. These include trade journals and unpublished research. Because the practitioner will be critically appraising the evidence, it is not essential for the research to have been peer reviewed.

Finding the best evidence in a feasible period of time calls for use of a computer with internet access to search online databases. There are many online databases focusing on various topic areas. The most useful for amplification evidence seem to be PubMed (<http://www.pubmed.gov/>), ComDisDome (<http://www.comdisdome.com/>), and CINAHL (<http://www.cinahl.com/>). PubMed is the most comprehensive, and it is free, but it is limited to peer-reviewed articles. ComDisDome lists dissertations and monographs that are widely read but not

Table 1. Elements of a Question for an EBP Critical Appraisal Activity

Element	Example
Person	Older woman who has occasional social contacts.
Problem	Moderate bilateral presbycusis and difficulty understanding speech in a social setting.
Proposed treatment	Digital hearing aids.
Comparison treatment	Programmable analog hearing aids with similar gain and input/output specifications.
Data	Speech intelligibility in babble—objective and/or subjective.

peer-reviewed, CINAHL lists some trade journals. Access to some databases requires a subscription. Libraries typically have subscriptions that allow searching many databases. Each database has slightly different rules for searching, and they always provide guidelines for the searching rules. It is well worth investing a little time to make sure you know how to search the database efficiently.

Returning to the question shown in Table 1, we will review a search of the PubMed database for evidence on this question. The following search items were typed into the search field: “hearing aid AND digital AND (analog OR analogue) NOT (implant OR implantable).” PubMed allows you to limit, or refine, the search in several ways that help weed out irrelevant references. Limits used for this search were as follows: age = “Adult,” language = “English,” publication date = “From 1995 To 2004,” and the box “only items with abstracts” was selected. This search yielded a list of 13 recent English-language articles on the topic of comparing analog and digital hearing aids in adult patients. The next step is to perform a preliminary review for direct relevance, based on the article titles. This review resulted in elimination of five articles with eight articles remaining. Thus, in less than five minutes, this search yielded a list of eight sources of potential evidence on the immediate problem facing

Mrs. Tribbet and her family. Table 2 lists the articles retrieved by this search.

Step 3: Evaluate the Evidence

The next step calls for the practitioner to critically appraise the evidence resulting from the search. It is useful to begin by reading the abstract for each article. This is provided as part of the reference in the online database. Begin with the most recent article and work back in time. Use the abstracts to do a preliminary evaluation of possible strengths and weaknesses of the reported data. Methods for evaluating strengths and weaknesses of data are discussed later in this article. Based on your evaluation of abstracts, select for further appraisal several articles that appear to provide the strongest evidence. Eliminate the rest from consideration at this time (although you might return to them later). How many articles should you plan to assess? If you have been fortunate enough to find a well-done systematic review of your topic (see later for systematic reviews), this might be all you will need. Often, there is no up-to-date systematic review. In this case, it is possible to look at four to five primary sources of evidence in a reasonable period of time.

At this stage, it is necessary to get the full text of each selected article. Because there are relatively few publications that feature

Table 2. Results of a Simple Search of the PubMed Database for Evidence on the Question in Table 1

Authors	Title	Reference
1: Wood SA, Lutman ME.	Relative benefits of linear analogue and advanced digital hearing aids.	<i>Int J Audiol.</i> 2004 Mar.; 43(3):144–155.
2: Parving A, Sibelle P.	Clinical study of hearing instruments: a cross-sectional longitudinal audit based on consumer experiences.	<i>Audiology.</i> 2001 Jan.–Feb.; 40(1):43–53.
3: Walden BE, Surr RK, Cord MT, Edwards B, Olson L.	Comparison of benefits provided by different hearing aid technologies.	<i>J Am Acad Audiol.</i> 2000 Nov.–Dec.; 11(10):540–560.
4: Arlinger S, Billermark E.	One year follow-up of users of a digital hearing aid.	<i>Br J Audiol.</i> 1999 Aug.; 33(4):223–232.
5: Bille M, Jensen AM, Kjaerbol E, Vesterager V, Sibelle P, Nielsen H.	Clinical study of a digital vs. an analogue hearing aid.	<i>Scand Audiol.</i> 1999; 28(2):127–135.
6: Boymans M, Dreschler WA, Schoneveld P, Verschuure H.	Clinical evaluation of a full-digital in-the-ear hearing instrument.	<i>Audiology.</i> 1999 Mar.–Apr.; 38(2):99–108.
7: Valente M, Fabry DA, Potts LG, Sandlin RE.	Comparing the performance of the Widex SENSO digital hearing aid with analog hearing aids.	<i>J Am Acad Audiol.</i> 1998 Oct.; 9(5):342–360.
8: Arlinger S, Billermark E, Oberg M, Lunner T, Hellgren J.	Clinical trial of a digital hearing aid.	<i>Scand Audiol.</i> 1998; 27(1):51–61.

amplification articles, it is likely that you will have subscriptions to some of the journals containing articles you need. Other articles can be obtained from libraries or online. The main trade journals are typically freely available online. Peer-reviewed journals can be obtained directly from the publisher or through services such as Ingenta (<http://www.ingentaconnect.com/>). The database usually shows how the full text of the article can be obtained. Read each article and assess the strengths and weaknesses of the evidence presented (see later for guidelines). If you decide you need to seek more evidence, you might return to the original list of articles or search article reference lists for additional leads.

Step 4: Recommendation

As you familiarize yourself with the evidence, it is important to consider the similarities and differences between the research subjects and your patient. Factors that should be considered include age, health, education, socioeconomic conditions, lifestyle, gender, and so on. These deliberations will help you determine the extent to which the available evidence applies to your patient. This is where “considered judgment” comes into play. Considered judgment is the process in which the practitioner combines his/her clinical experience and insights about the patient’s unique predicament with knowledge of the evidence derived from the critical appraisal. After weighing these matters, the practitioner decides on the appropriate recommendation for this patient and discusses the recommended plan with the patient.

Step 5: Follow-Up

The final step is often overlooked in practice, but it should not be. As the recommendation is presented to the patient, it is important to formulate a specific plan for reviewing the success of the recommendation. This could be a follow-up clinic visit in two weeks, a mailed questionnaire, a telephone call, et cetera. This step provides the opportunity for the practitioner to refine his or her skills in this particular problem area. Equally important, this step provides an opportunity to modify the recommendation if it is not successful for the patient. The result

adds to the material that can be brought to considered judgment the next time a similar problem is encountered.

HOW DO PRACTITIONERS EVALUATE EVIDENCE?

Undertaking the responsibility for critically evaluating evidence is somewhat daunting for many practitioners. Most current day hearing-care providers have minimal exposure to the principles involved in assessing research strengths and weaknesses. In the traditional mode of practice, this has been considered the province of research-oriented academics. However, in the current practice environment, with the accelerating rate of new knowledge production, it is insufficient for a practitioner to rely entirely on others to assess, distill, and impart information about the effectiveness of treatments for hearing loss. This is why the emphasis on EBP is growing. Fortunately, useful principles for critiquing evidence can be presented rather expeditiously.

At the outset, it is important to note that there are different types of research in the amplification literature, and not all of it reflects on effectiveness of treatments. There is a substantial body of research that has produced tutorial data. These kinds of studies are indispensable for teaching practitioners *how to do the job* of providing amplification most proficiently. A good example of this kind of research is found in Hawkins and Cook (2003). This work explored the extent to which simulated insertion gain provided by hearing aid fitting software was an accurate prediction of the real ear insertion gain observed on individual patients. The results produced valuable guidelines for practitioners about how to use and interpret simulated insertion gain estimates. Tutorial research plays an essential role in monitoring and enhancing the quality of treatment applications. One of the most important functions of textbooks and professors is to make sure that practitioners are aware of these types of data. Because it does not assess treatment effectiveness, tutorial research will not usually be critically appraised in an EPB search.

The type of research that provides information about the effectiveness of treatments typically involves investigations in which the treatment, such as a new type

of hearing aid technology, is used and evaluated on real people and perhaps compared with other potentially useful treatments. These studies provide information about *what job to do*, that is, what works best. This type of research is the focus of EBP critical appraisals. A recent example of effectiveness research was a trial of compression processing strategies reported by Larson et al (2000). This research asked which of three types of sound processing was most successful in hearing aids fitted to individuals with certain audiogram features. With the current rapid pace of technology advances and application developments, textbooks and experts cannot be relied on for comprehensive state-of-the-art knowledge about the effectiveness of all available treatments for a particular problem. These information sources are often outdated or they might not consider all available evidence on a topic.

Evaluating evidence after you find it involves determining the answers to three questions: How true or trustworthy is the evidence? How important or valuable is the treatment compared with other potential treatments? How relevant or applicable is the evidence to your patient? We will discuss how to approach each of these questions.

HOW TRUE OR TRUSTWORTHY IS THE EVIDENCE?

Evidence is produced in many different ways, using different approaches to research design. Inherent aspects of diverse research designs make them more or less resistant to weaknesses. These weaknesses could allow the data to be influenced by variables other than the treatment under evaluation. When this occurs, the data are not fully valid. In addition, regardless of the intended research design, an investigation might be compromised by any of several flaws. Depending on the strengths of the research design and the inherent or unavoidable limitations of the study, some types of evidence data are considered more valid than others and thus more likely to represent the reality of the treatment's effectiveness. The first job in critical appraisal of a source of evidence is to assess the design of the research and the strengths and weaknesses of the study.

Based on the results of this assessment, each source of evidence is assigned a "level

of evidence" score on a scale from 1 (highest level, most trustworthy) to 6 (lowest level, least trustworthy). If the research design is well executed, that is, relatively free of problems that might cause it to be invalid, the study is "high quality." For high quality studies, there is general agreement on the ranking of types of evidence from highest to lowest. Higher level evidence gives stronger support for your treatment recommendation and allows you to be more confident about the appropriateness of the recommendation for your patient. Levels of evidence are discussed in more detail below.

Assessing the Research Design

The evidence supporting hearing health care treatments is most frequently produced by one of the following: randomized controlled trial, nonrandomized intervention study, nonintervention descriptive study, case study, and expert armchair opinion. Each one will be briefly described.

Randomized Controlled Trial

The fundamental characteristic of a randomized controlled trial (RCT) is that the subjects are identified first and then randomly allocated into two or more experimental groups by a process equivalent to rolling a die. One group will receive the treatment of interest (new hearing aid, group counseling program, etc.). The other group will receive no treatment or a standard comparison treatment. Random allocation of subjects to groups tends to make the groups equal on all other variables (but this is only achieved in reality when a sufficient number of subjects is used). Once the groups are equalized by randomization, the only difference between the groups is the treatment variable. Except for the treatment variable, the groups are managed identically throughout the study. To promote equivalent treatment of all subjects, no matter which group they are assigned to, both researchers and subjects should be "blinded" about (i.e., unaware of) the group assignment of the subjects.

A valuable variant of the RCT design is a crossover design. This type of study begins like an RCT in that the subjects are first identified and then randomized into two (or more) treatment groups. Each group receives

a different treatment (e.g., Group 1 wears linear hearing aids, and Group 2 wears compression hearing aids). Then, after experiencing the assigned treatment for a specified period, each subject “crosses over” and receives the other treatment for a period of time (e.g., Group 1 now wears compression hearing aids while Group 2 wears linear hearing aids). In a crossover design, all subjects eventually receive, and provide data for, all of the treatments. So there are no questions about group equivalence when the treatments are compared with each other. In RCT and crossover designs, the validity (truthfulness) of the data is maximized because the chance of confounding variables or bias is much reduced.

Nonrandomized Intervention Study

This type of design is sometimes called a quasi-experiment. It is an intervention study, that is, an experimental treatment is applied by the researcher for some of the subjects. However, this design does not use random assignment to create the experimental groups. Instead, the groups are formed on the basis of some a priori factor(s). For example, Group 1 might comprise patients with a diagnosed cochlear dead region, and Group 2 might comprise patients with similar audiograms but no cochlear dead region. A weakness of this type of design is that it is difficult to be sure that the groups are equal except for the presence or absence of a cochlear dead region. Typically, the design will call for matching the groups in terms of obvious potentially confounding variables such as audiogram thresholds, subject ages, and gender, in an attempt to control validity threats as much as possible.

Nonintervention Descriptive Designs

In these designs, no treatment or intervention is given by the researcher. Instead, variables are examined as they naturally occur for the purpose of describing outcomes, or comparing samples, or examining relationships among variables. There are three common subtypes of this design: the cohort study, the case-control study, and the cross-sectional survey.

Nonintervention Design 1: Cohort Study. A cohort study is one that attempts

to determine the effects of experience with a particular variable. In this design, outcomes for a group who have had the experience are compared with outcomes for another matched group from the same population who have no experience or a different experience. For example, Group 1 might be patients who experienced a professional hearing aid dispensing service, whereas Group 2 might be patients who chose to experience a mail-order hearing aid dispensing service. The researchers would compare the outcomes for the groups to assess the impact of the different dispensing service models. An important feature of the cohort design is that the exposure to the variable of interest is not under the researcher’s control: Subjects are allocated to the groups based on their preexisting experiences. This type of study may be prospective (running from the present into the future) or retrospective (beginning in the past and assessed in the present). There are numerous potential weaknesses in cohort studies that limit the ability of researchers to draw unequivocal conclusions about the effects of the variable under study. The groups might be rather different from each other on other important variables. For example, it is quite likely that individuals who choose the mail order option have a different economic background and perhaps different personality profile from those who elect the professional dispensary. We should also consider other potential confounders such as whether each group accurately represents the target population, whether experience with the variable has been assessed accurately (e.g., patients who buy a mail order hearing aid might previously have purchased one in a private practice setting), whether outcome assessment is blinded, and so on.

Nonintervention Design 2: Case-Control Study. A case-control study attempts to determine what factors or variables are predictive of a particular outcome. In this design, previous experiences or variables for a group who have a particular outcome are compared with previous experiences or variables for another matched group from same population who do not have the outcome. For example, Group 1 (cases) might be patients who are satisfied with their hearing aids, whereas Group 2 (controls) might be patients who are dissatisfied with their hearing aids. The researchers compare

the two groups in terms of different variables (such as cost of devices, patient personality, etc.) to see whether any of them seem to be predictive of the satisfaction outcome. The case-control study begins after the outcome has occurred. As with cohort designs, there are numerous potential weaknesses in case-control studies that limit the ability of researchers to draw unequivocal conclusions about the predictive value of specific variables. For example, it is important that both groups are drawn from the same target population, so inclusion/exclusion criteria must be the same for cases and controls. We also should consider issues such as percentage participation, that is, the percent of eligible patients in each group who agreed to participate in the study. If only a small percent of eligible patients agree to participate in one of the groups, there can be a problem generalizing results from those few subjects to the target population.

Nonintervention Design 3: Cross-Sectional Survey. In this type of study, a representative group of patients is examined, tested, interviewed, et cetera, to determine both their treatment precursors and treatment outcomes at the same time. Statistical methods are used to explore the relationships between precursors and outcomes to determine what precursor variables might have influenced the outcomes. For example, precursor variables might be the hearing aid dispensing model (professional services or over-the-counter purchase), and the outcome variable might be satisfaction with the hearing aids. Cross-sectional surveys can use data from several types of sources such as questionnaires, case records, demographics, and so forth. Table 3 illustrates a hypothetical example of this type of study. Note that the composition of the groups that result from the survey cannot be predicted in advance and are often quite different in size. It is generally impossible to ensure that potential confounding variables are equally distributed among these groups. Because cross-sectional surveys can establish associations between variables, they can

provide noteworthy insights and provoke valuable hypotheses for future research. However, these studies cannot prove that any particular precursor variable caused a specific outcome to occur. This mistake in interpretation is frequently seen.

Case Report

This type of evidence involves detailed description of the history of a single patient, or a series of patients with similar problems. The case report can produce useful insights and hypotheses by illustrating previously unnoticed features of the condition, or the outcomes of specific treatments in these cases. An interesting example of a case report series is found in Carter et al (2001). These authors described four older patients with bilaterally symmetrical hearing loss who all exhibited low scores on a dichotic digits test and all preferred unilateral amplification over bilateral. This type of careful observation does not support definitive conclusions. However, it can stimulate research that might produce valuable evidence about the effectiveness of treatments for particular patients.

Expert Opinion (without Data)

This type of evidence is an untested point of view that may be drawn from a knowledge of basic disorders or systems, or extensive clinical experience, or logical deduction, or a combination of these. After numerous repetitions, expert opinion may become accepted as conventional wisdom. An example of widely encountered expert opinion is the assertion that wide dynamic range compression (WDRC) with a very low input compression threshold is the best type of hearing aid processing for individuals with outer hair cell (OHC) loss. The rationale is that OHC loss results in the absence of the cochlea's compressive function, and low-threshold WDRC compensates for this.

Table 3. Hypothetical Example of a Cross-Sectional Survey Involving 100 Subjects

	Professional dispensing services	Over-the-counter purchase
Satisfied patients	42	32
Dissatisfied patients	18	8

Note: Each cell shows the number of subjects found to have each combination of precursor (dispensing) and outcome (satisfaction) variables.

Assessing the Research Limitations

As noted earlier, to determine the level of evidence for a particular study, you must weigh the strengths and weaknesses of the research to assess its quality. If the research has substantial weaknesses, its level must be downgraded. Common weaknesses in amplification research include lack of randomization, not accounting for dropouts, absence of blinding, too few subjects, unrepresentative sampling, subject self-selection, and use of surrogate endpoints. Each of these will be briefly discussed.

Lack of Randomization

If the design does not include randomization of subjects to groups, it is likely that the experimental groups will not be equal at the outset. Therefore, group differences that appear in the outcomes of the study cannot be confidently ascribed to the effects of the treatments. For example, without randomization, it might be tempting to choose subjects for the experimental group who seem the most likely to benefit from the treatment. In this case, it is likely that the positive effects of the treatment will be exaggerated. To avoid this kind of problem, subjects should be assigned to groups by a method equivalent to tossing a coin.

Dropouts Not Accounted For

Ideally, the results for all the original subjects should be analyzed at the end, even if they do not complete the study. This is the basic premise of "Intention To Treat" analyses that are discussed in medical textbooks. This type of analysis is sometimes not possible in amplification research because subjects who drop out of the study may be lost to follow-up. However, as a minimum it is important to report how many subjects started, how many finished, and why some dropped out. This information can be essential for appropriate interpretation of the study results. For example, consider what happens if half the subjects in the group wearing new technology devices drop out of the study because they find the hearing aids unbearably uncomfortable. If the outcomes are analyzed using only data from the remaining subjects, the result will be deceptively positive. In any study, if a sizable percentage of subjects drop

out and are not accounted for in a plausible way that is not related to the outcome, the validity of the conclusions must be questioned.

Absence of Blinding

Blinding is the process of concealing a subject's group assignment from the subjects and/or the researchers. In single blinding, only the subjects or only the researchers know which group a subject is assigned to. In double blinding, neither subjects nor the researchers who collect the data know this information. Without blinding, there is a tendency for both subjects and researchers to be unwittingly influenced by a halo effect that suggests that the new treatment is better. Blinding is essential if subjective judgments such as questionnaires or rating scales are used as outcome data. Blinding also may be important for many so-called objective tests such as word repetition tests that are scored by listening to spoken responses. Sometimes blinding is difficult or impossible, as when bilateral hearing aid fitting is compared with unilateral hearing aid fitting. Nevertheless, lack of blinding is still a limitation that can affect the validity of the results of the study.

Too Few Subjects

Many studies of amplification reported in peer-reviewed journals are underpowered. That is, they did not use enough subjects to detect a clinically worthwhile treatment effect or the difference between two treatments. If a study does not produce a statistically significant result, this is often interpreted to mean that no interesting effects or treatment differences existed. If the study was underpowered, this interpretation can be completely wrong and misleading. Therefore, it is important to consider the power of a study before accepting the truthfulness of null results. The number of subjects needed to give adequate power to any particular study is determined by the smallest interesting effect (for example, the smallest difference in effectiveness between two technologies that would be of practical clinical value) and the inherent variability of the test instrument (such as a speech recognition test or a questionnaire used to evaluate the two technologies). Jones et al (2002) provided a revealing review of the frequent inadequacy

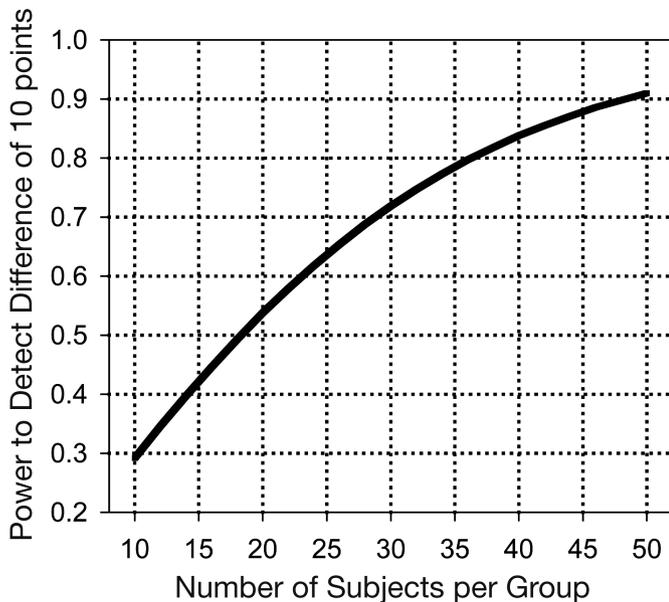


Figure 1. The power of an experiment to detect a difference as small as 10 points, as a function of the number of subjects used in each group, assuming a within-group standard deviation of 15 points.

of experimental power in stuttering research. Their observations apply equally to many published studies of amplification.

Pressure is intensifying to require explicit justification of sample size in all scientific publications. This would not be an unreasonable requirement. Although power computations were regarded as arcane not long ago, software for this purpose is now widely available and simple to use. See, for example, the software available at <http://www.stat.uiowa.edu/~rlenth/Power/>. The principles of power calculation are illustrated in Figure 1. To generate this fictional illustration, it was assumed that the experimental design was an RCT involving a comparison of two hearing aids with different technologies. Each technology was worn by a different group of subjects. The technologies were evaluated using a speech recognition test that was known to have a between-subject standard deviation of 15 points. It was assumed that either hearing aid might be superior and that the smallest interesting difference between them was 10 points; that is, a difference smaller than 10 points was of no practical consequence. Figure 1 shows the power of the experiment to detect

a difference as small as 10 points, as a function of the number of subjects used in each group. For example, if each group had 15 subjects, the experiment would have a power of about 0.4. This means that, if there is a difference of 10 points between the two hearing aids, about 40% of studies with 15 subjects in each group would be expected to yield a result that was statistically significant at the $p < .05$ level. Conversely, even when a true difference of 10 points exists between the two hearing aids, 60% of studies with 15 subjects in each group will not detect this difference by producing a statistically significant result. If the number of subjects in each group is increased to 50, Figure 1 shows that the power of the experiment will increase to about 0.9. Thus, it is very likely that a difference of 10 points would be detected in this experiment. Most researchers agree that a power of about 0.8 is a reasonable goal in study design. Based on this guideline, our comparison of the two hearing aids should include groups of about 37 subjects each.

Inadequate Sampling

Unless the method of selecting and recruiting subjects is carefully planned, it can happen that the people who are recruited for the research may not be representative of the population of interest. This might be called experimenter selection bias. For example, a laboratory might maintain an ongoing group of subjects who have been in many studies. These individuals are not necessarily representative of the general population of hearing aid wearers for the variables of interest in a particular experiment. Similarly, a newspaper advertisement may interest certain types of people more than others. Problems also can occur when all the subjects are recruited from a single source such as one affluent retirement community, or a hospital that serves only male veterans. Regardless of the outcome of the study, it is important to consider the extent to which it is valid to generalize the results.

Subject Self-Selection Bias

This problem is related to the sampling issue discussed above. The researchers might make substantial efforts to recruit subjects who are widely representative of the

population of interest, but individuals who fall into certain categories might tend to decline to take part in the experiment. The “participation rate” is defined as the number of study volunteers divided by the total number of eligible subjects asked to participate. For example, in a study of reasons for hearing aid returns, perhaps 80% of those who kept the hearing aids agreed to participate when asked, but only 10% of those who returned the hearing aids agreed to participate. Researchers must bear in mind that volunteers might be different from those who refuse. If the participation rate is low or is much different for each group, this suggests a problem with self-selection bias. To allow critical appraisal of the research, participation rate should be reported separately for each group.

Use of Surrogate Endpoints

A surrogate endpoint is an intermediate variable that is used as an outcome variable and assumed to predict the real outcome of interest. Surrogate endpoints that have been used in amplification research include insertion gain and monosyllabic word recognition. Usually, these are not intrinsically important outcomes, but they are assumed to be predictive of the important outcomes. A surrogate endpoint is only valid to the extent that it has a known relationship to real outcome of interest. For many studies of amplification, the only really important outcome is what happens in daily life.

There are very few research endeavors that are completely free of flaws. These limitations often occur because restricted resources make compromises unavoidable. Nevertheless, it is important for researchers to recognize the weaknesses that might limit the validity of study outcomes. Weaknesses

should be made explicit in presentation of the results so that professional consumers can take account of those concerns in the process of critical appraisal.

Assigning a Level of Evidence to Each Study

The final decision about the validity of the evidence produced by a particular study is made on the basis of a consideration of the inherent strengths of the study design and any weaknesses that diminished the quality of its execution. This decision leads to the assignment of a level of evidence for the study. The levels of evidence typically span a scale from 1 to 6, with 1 being highest, or most trustworthy, and 6 being lowest, or least trustworthy. Several level-of-evidence hierarchies have been suggested and can be found online. They differ very slightly, based on the field of research they are oriented toward. Table 4 illustrates a hierarchy of evidence levels that applies well to the field of amplification research (adapted from Harbour and Miller, 2001).

Note that Level 1 evidence is reserved for systematic reviews and meta-analyses. Both of these are approaches to combining evidence from several studies. A systematic review identifies, evaluates, and synthesizes studies of a selected topic (such as the effectiveness of group hearing therapy) using explicit but largely qualitative rules. A meta-analysis is a quantitative approach to combining and summarizing several studies of the same topic. Both approaches foster a global perspective that promotes an understanding of the weight of evidence on the topic of interest.

To be optimally useful, a systematic review or meta-analysis must be thorough and impartial as well as performed by

Table 4. Level of Evidence Hierarchy for High-Quality Studies

Level	Type of Evidence
1.	Systematic reviews and meta-analyses of randomized controlled trials or other high-quality studies.
2.	Randomized controlled trials.
3.	Nonrandomized intervention studies.
4.	Nonintervention studies: cohort studies, case-control studies, cross-sectional surveys.
5.	Case reports.
6.	Expert opinion.

Table 5. System for Quality Rating of Individual Studies

Rating	Interpretation of Rating
++	Very low risk of bias. Any weaknesses that are present are very unlikely to alter the conclusions of the study.
+	Low risk of bias. Identified weaknesses or omitted information probably would not alter the conclusions of the study.
—	High risk of bias. Identified weaknesses or omitted information are likely or very likely to alter the conclusions of the study.

individuals who are knowledgeable in both the topic area and research issues. Such individuals are more likely to identify the potential weaknesses in the summarized studies. A careful systematic review or meta-analysis of the effects of a particular treatment is a gold mine for the practitioner because it makes an additional critical review unnecessary. The amplification literature does not contain many of these kinds of endeavors as yet. However, several systematic reviews are included in this journal issue.

Assigning a level of evidence for a given study is quite straightforward in those cases where the research is essentially free from limitations or weaknesses. However, few studies rise to this lofty standard. Thus, it is desirable to elaborate the level of evidence table to account for variations in the quality of execution of the different research designs. Several approaches have been suggested to accomplish this. A simple but serviceable system is shown in Table 5. Quality ratings are usually not applied to case report or expert opinion types of evidence. Evidence levels 1, 2, 3, and 4 (Table 4) should have associated ++, +, or — grading from Table 5.

HOW IMPORTANT? DOES THE TREATMENT HAVE CLINICALLY SIGNIFICANT EFFECTS?

The discussion to this point has focused on determining how truthful or valid the evidence is. As noted earlier, the second component of critical appraisal calls for evaluating the extent to which the evidence indicates that the treatment is important. An important treatment is one that produces effects that are sufficiently large to be worth endorsing or including in one's practice. Sometimes this is called clinical significance (e.g., Jacobson and Truax, 1991; Thompson, 2002).

The traditional approach to assessing the importance of a research outcome is null hypothesis significance testing (NHST). NHST determines the likelihood that an observed result, such as a ten-point improvement in mean scores, was a chance occurrence. If it is determined that the likelihood of the result being due to chance is smaller than a criterion value, often 5%, then the NHST conclusion is that the result is statistically significant. Frequently, a statistically significant result is assumed to be an important result, but this is not necessarily the case. Even a very small improvement, say, one point, can be statistically significant, depending on the research design. However, a one-point improvement is not necessarily important to patients in the real world. Conversely, a nonsignificant statistical result is often assumed to be the equivalent of no effect, and this is also not necessarily true.

The problem with relying on NHST to evaluate effectiveness research is that the result of NHST is strongly affected by sample size: more subjects are more likely to produce significant results, even when the magnitude of the effect is actually too small to be important. Further, even when the effect is large enough to be interesting, an NHST result may be nonsignificant because the study is underpowered. These limitations ignore the reality that, for the individual patient, it is the magnitude of the effect that is important. As a result of these types of concerns, there is increasing pressure to include evaluations of *effect size* in treatment research. Effect size (ES) analyses have the advantages of focusing on the magnitude of the result and being independent of sample size. In addition, computing effect sizes promotes pooling of data across studies in meta-analyses. This allows us to combine several low-power studies to give a result

with greater power. Finally, when used with confidence intervals, effect size computations give the same information as NHST and more. Because of these advantages, journals are slowly changing their policies to stress the importance of reporting effect sizes when appropriate.

Effect size is a metric that expresses the magnitude of a result, such as a difference in mean scores for two different types of hearing aids, within the context of the expected individual variation in scores. Effect size can be computed by several different methods, depending on the type of data under analysis (see, for example, Hill and Thompson, 2004). A common approach is to compute ES as the standardized mean difference between the two groups (Cohen, 1988), as shown in the equation

$$ES = \frac{(\text{mean of experimental group}) - (\text{mean of control group})}{\text{pooled standard deviation}}$$

Figure 2 illustrates some features of effect size. Two pairs of curves (upper and lower boxes) depict hypothetical results (distribution of scores in each group) for two experiments that compared an old hearing aid with a new hearing aid. Let us call the upper box "Study A" and the lower box "Study B." Note that the mean difference between the old hearing aid group and the new hearing

aid group was the same in both studies, and assumed to be 17 points. However, the variability of scores for the groups is larger in Study B (standard deviation = 11) than in Study A (standard deviation = 4). Because the mean differences are equal while the variabilities are different, the overlap between the two groups is much less in Study A. This means that most of the subjects in Study A realized a better result with the new hearing aid, whereas some of the subjects in Study B probably had no improvement with the new hearing aid.

These two studies did not yield results with the same clinical significance. On the basis of this evidence, the practitioner could recommend the new hearing aid from Study A with a high level of confidence that it would produce an improvement for a given patient, whereas one could not be quite as confident about recommending the new hearing aid from Study B. Nevertheless, as long as the sample size for both studies was at least 10 subjects per group, both of these studies produced a statistically significant effect ($p < .05$) showing an improvement for the new hearing aid. Therefore, both studies yielded the same statistical (NHST) result. On the other hand, despite their NHST equivalence, the difference between the results of the two studies is accurately reflected in the calculated effect sizes. The effect size for Study A is 4.3, whereas the effect size for Study B is much smaller at 1.5. This example demonstrates a general principle, namely, a bigger effect size is more likely to indicate an important and clinically significant result.

Figure 3 illustrates another advantage associated with considering effect size. When the computed ES is combined with its confidence interval (CI), the results not only reveal the outcome of NHST but also allow us to determine whether the result is conclusive or not. The figure depicts results of four different studies comparing an experimental hearing aid with a comparison hearing aid. For each study, the effect size is shown (large dot), along with its associated 95% confidence interval (horizontal line). The 95% confidence interval is the range of effect sizes you would expect to measure for 95 out of 100 replications of the study.¹ If the confidence interval intersects the heavy vertical line showing an effect size of 0.0, the results of NHST for that study are not

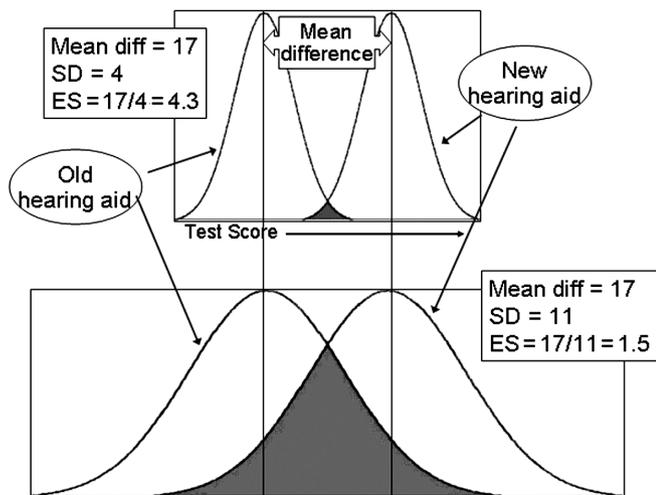


Figure 2. Hypothetical distributions of outcome scores in each group for two experiments that compared an old hearing aid with a new hearing aid. ES = effect size.

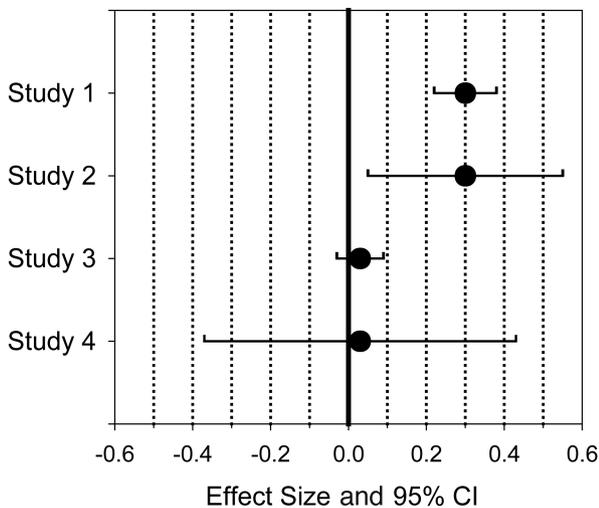


Figure 3. Results of four different studies comparing an experimental hearing aid with a comparison hearing aid. For each study, the effect size is shown (large dot), along with its associated 95% confidence interval (horizontal line).

statistically significant. If the confidence interval does not intersect the line showing an effect size of 0.0, the results of NHST for that study will be statistically significant ($p < .05$). Based on the ES and 95% confidence interval, each study supports a different conclusion.

- Study 1 produced an ES of 0.3 with a CI from 0.22 to 0.38. This CI does not intersect the 0 effect line, so we know that the study produced a statistically significant result. In addition, because the CI is relatively small, we can be confident that this result is highly reproducible. Therefore, the results of this study are conclusive: The experimental treatment yields a definite improvement over the comparison treatment, and the magnitude of the effect is very close to 0.3.
- Study 2 also produced an ES of 0.3, but the CI span was 0.05 to 0.55. This CI does not intersect the 0 effect line, so we know that this study also produced a statistically significant result. However, because the CI is relatively wide, we cannot be confident that the ES of 0.3 will be seen again if the study is replicated. In fact, we would expect to observe effect sizes ranging from very small (0.05) to as large as .55 in

subsequent studies. Therefore, the results of this study are inconclusive: it appears that the experimental treatment yields an improvement over the comparison treatment, but the magnitude of the effect might be too small to be important, or it might be rather large.

- Study 3 produced an ES of 0.03 with a CI from -0.03 to 0.09. This CI does intersect the 0 effect line, so we know that the study did not produce a statistically significant result. In addition, because the CI is relatively small, we can be confident that this outcome is highly reproducible. That is, additional studies will also yield a very small effect. Therefore, the results of this study are conclusive: The experimental treatment is no better or worse than the comparison treatment.
- Study 4 also produced an ES of 0.03, but the CI span was -0.37 to 0.43. This CI does intersect the 0 effect line, so we know that this study also produced a statistically nonsignificant result. However, because the CI is relatively wide, we cannot be confident that the ES of 0.03 will be seen again if the study is replicated. In fact, we would expect to observe effect sizes ranging from rather negative (-0.37) to as large as +0.43 in subsequent studies. Therefore, the results of this study are inconclusive: it has failed to rule out the presence of a potentially clinically important effect. The experimental treatment might actually be considerably more or considerably less effective than the comparison treatment.

Based on these interpretations, Studies 1 and 3 produced conclusive results whereas Studies 2 and 4 did not. From a pragmatic point of view, this way of depicting the data is much more informative about the likely importance of the evidence than a traditional NHST approach.

A final demonstration of the potential value of using effect sizes is shown in Figure 4. This figure illustrates the advantage of combining effect sizes from different studies to increase power. Data were interpolated from figures in two recent publications in which WDRC hearing aids were compared with compression limiting (CL) hearing aids.

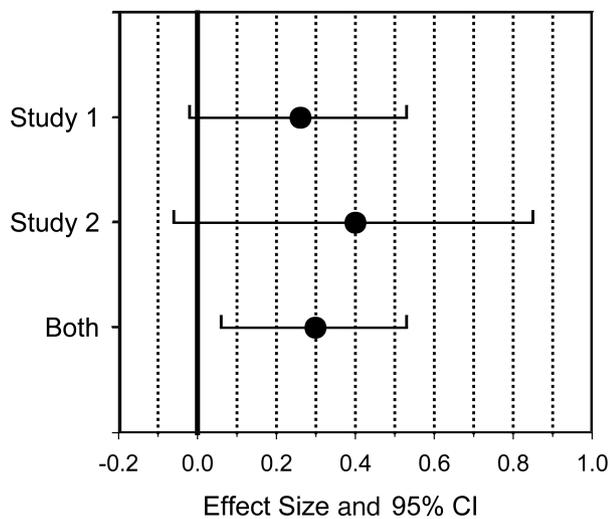


Figure 4. Illustration of combining effect sizes from two different studies to increase power.

Although neither study reported effect sizes, both included figures showing means and standard deviations. For Figure 4, mean scores were extracted from each study for understanding low-level speech in a background noise typical of living room listening. Effect sizes were then computed so that a positive effect indicated better performance with the WDRC processing. As Figure 4 shows, Study 1 yielded an ES of 0.26, and Study 2 yielded an ES of 0.4. For both studies, the computed 95% CI intersected the zero effect line, implying that WDRC processing was not superior to CL processing. This was consistent with the NHST result reported in both studies. Commercially available software was used to combine the results from the two studies in a small scope meta-analysis.² The combined result, shown in Figure 4, produced a combined ES of 0.3 and a CI that did not include 0, indicating that WDRC processing was indeed reliably superior to CL processing for this listening condition. Note that the result is still somewhat inconclusive in that the CI spans a range from a very small effect (0.06) to a moderately large effect (0.53). However, continued careful addition of data from other studies to the meta-analysis would probably reduce the uncertainty of the outcome.

How big does the effect size need to be in

order to be important? This question cannot be answered insightfully without considering the specific study in the context of the effect sizes that have been found in other related types of studies. Although some authors have tentatively suggested benchmarks for small, medium, and large effects, it is widely recognized that whether an effect is regarded as consequential or trivial depends on the specific context. For example, an effect equal to one life saved could be important, whereas one more vitamin pill consumed might be trivial. One approach to considering the importance of effect size in amplification research might be to consider the overlap between experimental groups (the shaded areas in Figure 2) that is associated with different effect sizes. If the ES is 2.0, there is only about 20% overlap between the treated and control groups, indicating that most of the patients in the treatment group performed quite differently from those in the control group. If the ES = 0.3, there is about 80% overlap between these groups, indicating that most of the patients in the treatment group did not perform differently from the control patients. At present, relatively few studies of amplification effectiveness have reported effect sizes. Including this type of information in published reports should become a high priority for the research community, so that the context needed to evaluate new treatments will be available.

In summary, evaluation of the importance of evidence calls for more than traditional significance testing. Effect sizes and associated confidence intervals are essential information for practitioners in judging the clinical significance of evidence about the effectiveness of treatments, their superiority compared to a standard treatment, and the likelihood that each individual patient will participate in the improvements offered by the treatment. This type of information ultimately will allow practitioners to determine whether they can be confident about recommending new innovations, or whether these should be adopted with caution, if at all.

HOW RELEVANT? DOES THE EVIDENCE APPLY TO MY PATIENT?

The third and final component of critical appraisal calls for evaluation of the extent to which the evidence applies to the

practitioner's patient(s). This involves considering whether the study subjects were similar to the patients on variables such as age, financial situation, health, living arrangement, lifestyle, and so forth. The most relevant evidence is derived from studies that used subjects who were similar to clinical patients on the pertinent variables.

It is also important to determine whether the study was an efficacy study or an effectiveness study. Definitions of efficacy and effectiveness were first proposed by Cochrane (1972). An efficacy study asks whether the treatment *can work* under optimal conditions. These studies are intentionally designed to be carried out under the best possible circumstances. Usually, the subjects are highly motivated and have relatively uncomplicated hearing impairments. The involved practitioners are extensively trained in using the treatment and have adequate time to make sure that the application is fully correct. In addition, efficacy studies often employ lengthy test regimes that would not be suitable for routine clinical use.

An effectiveness study asks whether the treatment *does work* under typical conditions. In these studies, subject motivation is often complicated by issues such as other health problems, pressures of everyday living, financial concerns, personal crises, vacation schedules, and so on. Practitioners involved in the study may have minimal training in using the treatment and limited time to refine its application, as well as the distractions and pressures of everyday practice. Data collected in effectiveness studies should always include a real-life field trial. Although efficacy studies are essential in the research and development phase of a new treatment, it is quite possible for a treatment to be efficacious without being effective. Evidence of real-world effectiveness is more relevant to your patient than evidence of ideal-world efficacy.

It is worth noting that Cochrane (1972) also defined a third type of study—an efficiency study. An efficiency study asks how well the treatment works in comparison to what it costs in dollars and other resources. In other words, is it cost-effective? This is also an important type of question for amplification researchers and practitioners to consider. So far, there are relatively few such studies, but see, for example, Taylor et

al (2001), Abrams et al (2002), and Joore et al (2003).

THE RECOMMENDATION: COMBINING TRUTHFULNESS, IMPORTANCE, AND RELEVANCE

Your appraisal of truthfulness, importance, and relevance of evidence is unified into a recommendation through the use of considered judgment. This is the point in the process where the practitioner's clinical experience and skill, and understanding of the patient's needs and priorities, is blended with knowledge of the strengths, weaknesses, and applicability of the evidence, and the clinical significance of the treatment under consideration. Figure 5 schematically illustrates the application of the EBP method to generate a recommendation.

In each individual case, considered judgment will yield the best recommendation possible, based on your knowledge of the patient and the evidence. It will not always be possible to base your recommendation on high-level evidence (Levels 1 or 2) because relatively few studies that fulfill these criteria are presently available in the amplification literature. Nevertheless, recommendations must be made, and so compromises must be accepted. This does not invalidate the process. At the same time, because the evidence may be less than totally convincing, it is essential for the practitioner to recognize the degree of confidence you can have that your best recommendation will be useful and appropriate for your patient. This calls for you to determine the grade of the recommendation.

The grade of the recommendation is an indicator of the extent to which the weight of the evidence supports the recommendation. Table 6 gives a system for grading recommendations that is applicable to amplification research and is generally consistent with similar systems that have been suggested (e.g., Harbour and Miller, 2001). A recommendation with a grade of "A" is one that the practitioner can feel very confident in making because it is supported by a body of high-quality, relevant data and is therefore very likely to be helpful and appropriate for the patient. On the other hand, a grade "D" recommendation must be made with considerable caution because its evidence base is low quality, which means

Table 6. System for Grading a Recommendation

Grade	Criteria for grade assignment
A	Level 1 or Level 2 studies with consistent conclusions.
B	Consistent Level 3 or 4 studies or extrapolated evidence* from Level 1 or 2 studies.
C	Level 5 studies or extrapolated evidence from Level 3 or 4 studies.
D	Level 6 evidence or inconsistent or inconclusive studies of any level or any studies that have a high risk of bias.

Note: It is assumed that the level of evidence assigned to a study has been based on both the study design and the quality of its execution.

*Extrapolated evidence is evidence that is generalized to a situation in which it is not fully relevant, for example, when results obtained with young adults are used to support a recommendation for an older person.

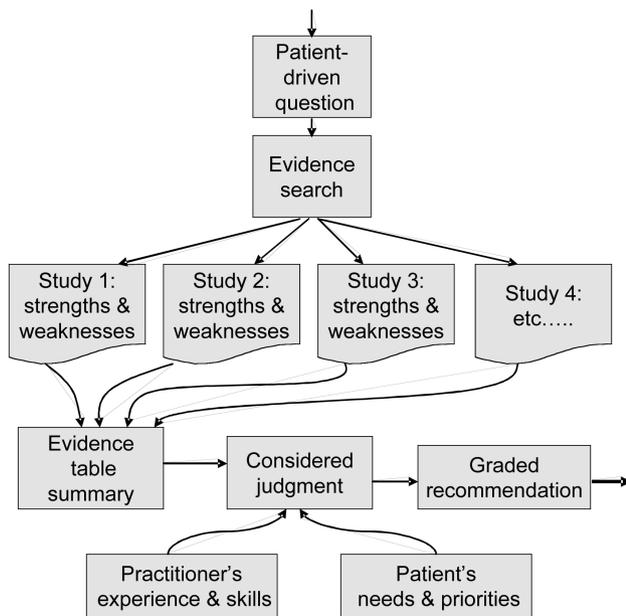


Figure 5. Schematic illustration of the application of the EBP method to generate a recommendation.

that it might not be very helpful or appropriate for the patient. Reviews of the amplification literature suggest that there is currently relatively little evidence that would support grade “A” recommendations. However, it is realistic to expect the research community to produce evidence that would support grade “B” recommendations.

The grade of the recommendation provides a basis for expectation about the outcome of the course of action proposed by the practitioner. Many practitioners feel that these expectations should be shared with the patient in order to enlist the patient as a partner in the rehabilitation process and to reassure the patient that you are following state-of-the-art practices with a strong

scientific base. It is also essential to plan an explicit method for evaluating the outcome of your recommendation (EBP Step 5). This will generate knowledge that can be brought to bear in the considered judgment phase in the future.

WHAT CAN RESEARCHERS DO TO SUPPORT EVIDENCE-BASED PRACTITIONERS?

Despite its widespread adoption in other health-related professions, implementation of evidence-based principles is not frequently seen in amplification practices at this time. This is partly because many practitioners are not familiar with the methods of EBP and critical appraisal of evidence. This can be remedied through continuing education forums and inclusion in academic curricula. However, it is also imperative that amplification researchers pay close attention to the needs of practitioners when generating and disseminating evidence. There are at least three components of research endeavor that are needed. First, individual studies must be well designed with careful consideration of validity threats such as halo effect, inadequate power, and surrogate outcomes. Second, there is a desperate need for published systematic reviews and meta-analyses of related investigations that currently exist in the literature. These types of articles will be enormously helpful to practitioners searching for evidence, as well as necessary precursors to developing systematic research programs that will rapidly move the field forward. The ability to produce meta-analyses will be facilitated if researchers provide data for effect sizes or, at least, data that allow them to be computed (such as means and standard deviations).

Third, existing evidence must be readily accessible to practitioners. This calls for meticulous, complete reporting that promotes the critical appraisal task. Whenever possible, the evidence should be published in an indexed location. Key words should be carefully considered so that the article will be detected when appropriate in an online database search. The title should explicitly describe the research so that the article will survive an initial title-based screening for content. In addition, a structured abstract considerably facilitates preliminary screening of the article for applicability to the problem at hand.

FINAL COMMENT

Moving from the traditional practice mode to the evidence-based practice mode will require commitment, perseverance, and patience. Practical and psychological obstacles will be encountered that add to the difficulty of making fundamental changes. Practitioners may feel challenged because they have not been prepared to carry out the searching and appraisal activities required by EBP. We can expect to see an increase in professional learning opportunities to address this concern. EBP might seem inefficient or not viable because of the time consumed in searching out and appraising relevant sources. One approach to managing this is to limit EBP searches to no more than one new problem per week. In addition, searches take less time as practitioners become more proficient with practice. Practitioners might become frustrated when high-quality evidence is lacking. Keep in mind that the first step toward improving the evidence is a recognition of the necessity for improvement. Those who are responsible for making rehabilitation recommendations should expect and demand appropriate evidence to support those recommendations.

NOTES

1. See Cumming and Finch (2001) for details about computation of confidence intervals for effect sizes.
2. Exploratory Software for Confidence Intervals. Accessed at www.latrobe.edu.au/psy/esci on March 4, 2003.

REFERENCES

- Abrams H, Chisolm TH, McArdle MS. (2002) A cost-utility analysis of adult group audiologic rehabilitation: are the benefits worth the cost? *J Rehabil Res Dev* 39:549–558.
- American Speech-Language-Hearing Association. (2005) Evidence-Based Practice in Communication Disorders [Position statement]. http://www.asha.org/NR/rdonlyres/4837FDFC-576B-4D84-BDD6-8BFF2A803AD3/0/v4PS_EBP.pdf
- Amlani A. (2001) Efficacy of directional microphone hearing aids: a meta-analytic perspective. *J Am Acad Audiol* 12:202–214.
- Carter AS, Noe CM, Wilson RH. (2001) Listeners who prefer monaural to binaural hearing aids. *J Am Acad Audiol* 12:261–272.
- Cochrane AL. (1972) *Effectiveness and Efficiency: Random Reflection on Health Services*. London: Nuffield Provincial Hospitals Trust.
- Cohen J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- Cumming G, Finch S. (2001) A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educ Psychol Meas* 61:532–574.
- Dollaghan C. (2004) Evidence-based practice: myths and realities. *ASHA Leader* (Apr. 13):4–5, 12.
- Hanratty B, Lawlor D. (2000) Effective management of the elderly hearing impaired—a review. *J Public Health Med* 22:512–517.
- Harbour R, Miller J. (2001) A new system for grading recommendations in evidence based guidelines. *BMJ* 323:334–336.
- Hawkins DB, Cook JA. (2003) Hearing aid software predictive gain values: how accurate are they? *Hear J* 56:26–34.
- Hill CR, Thompson B. (2004) Computing and interpreting effect sizes. In: Smart JC, ed. Vol. 19 of *Higher Education: Handbook of Theory and Research*. Boston: Kluwer, 175–195.
- Jacobson NS, Truax P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 39:12–19.
- Jones M, Gebiski V, Onslow M, Packman A. (2002) Statistical power in stuttering research: a tutorial. *J Speech Lang Hear Res* 45:243–255.
- Joore MA, Van Der Stel H, Peters HJ, Boas GM, Anteunis LJ. (2003) The cost-effectiveness of hearing-aid fitting in the Netherlands. *Arch Otolaryngol Head Neck Surg* 129(3):297–304.
- Kochkin S. (2001) The VA and direct mail sales spark growth in hearing aid market. *Hear Rev* 8(12):16–24, 63–65.
- Larson VD, Williams DW, Henderson WG, Luethke LE, Beck LB, Noffsinger D, Wilson RH, Dobie RA, Haskell GB, Bratt GW, Shanks JE, Stelmachowicz P,

Studebaker GA, Boysen AE, Donahue A, Canalis R, Fausti SA, Rappaport BZ. (2000) Efficacy of 3 commonly used hearing aid circuits: a crossover trial. *JAMA* 284(14):1806–1813.

Maki-Torkko E, Brorsson B, Davis A, Mair L, Myhre K, Roine R, Rosenhall U, Sorri M, Stilven S. (2001) Hearing impairment among adults—extent of the problem and scientific evidence on the outcome of hearing aid rehabilitation. *Scand J Audiol* 54:8–15.

Medwetsky L, Sanderson D, Young D. (1999) A national survey of audiology clinical practices, part 2. *Hear Rev* 6(12):14–22.

Robinson K. (1999) Evidence-based medicine and its implications for audiological science. *Br J Audiol* 33:9–16.

Sherbourne K, White L, Fortnuni H. (2002) Intensive rehabilitation programmes for deafened men and women: an evaluation study. *Int J Audiol* 41:195–201.

Taylor R, Paisley S, Davis A. (2001) Systematic review of the clinical and cost effectiveness of digital hearing aids. *Br J Audiol* 35:271–288.

Thompson B. (2002) "Statistical", "practical", and "clinical": how many kinds of significance do counselors need to consider? *J Couns Dev* 80(Winter):64–71.

Copyright of Journal of the American Academy of Audiology is the property of American Academy of Audiology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.