

# “Buzz Off, I Know What I’m Doing”: Protocols in EHDI

*Martyn Hyde*

## Background

The writing of this chapter was spurred by experiences of the author and several colleagues across Canada to the effect that efforts to encourage, explain and facilitate adherence to test protocols or guidelines in EHDI can be met with a wide range of responses from audiologists. The range has been from welcoming relief through indifference, scepticism, refusal and even outrage that anyone would try and tell other clinical professionals how to do their job. The title of this chapter reflects that experience.

Clinical protocols have been around for a very long time. One of the first recorded can be traced to around 2020 BC, in the reign of the god-king Umamma, in the third dynasty of the Mesopotamian city-state of Ur, located south east of modern-day Baghdad. An approximation to its wording is:

*“Take dried wine dregs, juniper and prunes. Pour on beer. Rub the diseased part with oil and bind on the mixture ...”*

Such a protocol would probably work for otitis externa but its effects for most other conditions are best left to the imagination. What is more interesting is that scepticism about clinical protocols, or at least concern about their effects and net benefit, also has a lengthy his-

tory. For example, around 400 BC no less a mind than Plato himself proposed a “thought experiment” that can be translated roughly as follows:

*“Suppose clinical practice were to be governed by rules that reflect the majority opinion of an expert group ... a concept that has merit... then given that the essence of effective clinical practice is flexibility and innovation ... would it not be endangered by such rules ...?”*

What is truly remarkable is that while clinical protocols and concerns about their appropriateness have been around for so long, the essence of the debate about their impact and merits has evolved rather little in its sophistication over a period of more than two thousand years. Plato’s statement could well have been taken from a 2010 publication addressing the impact of evidence-based guidelines, were it not for the elegance of his expression. These two snippets of ancient history encapsulate the central issue of this article – the case for adoption of strong clinical protocols and practice guidelines, in the context of programs for Early Hearing Detection and Intervention (EHDI).

## The Audiologist’s Challenge

Audiologists can have many roles in EHDI programs: supervision of newborn screening or even doing screening itself; diagnostic audiometry in newborns or infants who fail screening; informing and counselling families of infants with confirmed, permanent hearing loss (PHL); providing amplification, and support services for other devices such as cochlear implants; follow-up hearing assessments; surveillance of infants at high risk of PHL; and various other activities such as personnel training and program administration.

In each and every one of these activities, the central challenge to the audiologist is to provide the best possi-

---

**Address correspondence to:** Martyn Hyde, PhD., Suite 201, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. mhyde@mtsinai.on.ca

ble quality of care to each and every child and family. While it is not easy to carry out consistently good audiometry in adults, such work in infants is more demanding and especially so within an EHDI program context. For example, there are two major issues in EHDI involving pressure of time. First, the patient or client is a newborn or young infant, and they have their own agendas. They may or may not indulge the audiologist and even if they do cooperate for a while it may evaporate at any moment. This can have strong implications for effective and efficient diagnostic testing tactics and strategies, as will be discussed shortly. The second pressure of time arises because the whole point of EHDI is *early* detection and intervention. Unless timely and effective performance is achieved at every step along the care pathway through to language interventions, the whole point of newborn hearing screening is compromised.

The audiologist strives, of course, to do the best possible job by bringing to bear insights and skills acquired from training, clinical experience, knowledge of clinical and scientific literature, familiarity with professional practice guidelines and so on. However, in this day and age the volume and source diversity of potentially relevant information and evidence, whether good, bad or indifferent in their own quality, is overwhelming. It is more than a full time job even trying to keep pace with the flow of new data, let alone to evaluate it and make use of it in clinical practice. Protocols and guidelines are in part a response to this avalanche of publications and to the wide range of relevance, validity and generalizability reflected in primary clinical reports.

The audiologist's challenge is intensified because EHDI programs have been and continue to be subjected to what is arguably the closest scrutiny of any screening program in the history of public health. Ballooning healthcare costs, the pervasive zeitgeist of accountability, program evaluation and evidence-based healthcare, as well as the curious but familiar diversity in public and professional awareness of the importance of early language development, are all contributors to the challenging context in which EHDI programs must operate today. While EHDI programs have proliferated over the last decade, their survival and their lasting impact are questions driven ultimately by the quality of clinical services.

## What *is* Quality of Care?

One cannot possibly optimize anything rationally unless one is crystal clear about exactly what it is. There are many, many viewpoints about what service quality is.

It is different things to different stakeholders, such as funding decision-makers, program managers, clinical service providers and service recipients. But, a particularly clear, compelling and generalizable conceptual framework for it in the context of healthcare was conceived and promulgated by Avedis Donabedian (1919–2000), a polymath and celebrated authority on healthcare who ended his stellar academic career at the University of Michigan. Donabedian (1990) defined seven what he called “pillars” of quality; these are presented with very brief explanation (reflecting the present author's interpretation) as follows.

### Efficacy

*At its best, can it (i.e., whatever is being examined) improve health?* The emphasis here is on the word “can”. The *efficacy* of a procedure or process usually relates to whether it achieves its intended effect under ideal, restricted or “laboratory” conditions. For example, many randomized controlled trials of “treatment” efficacy impose strong constraints on subject eligibility, strong controls of how the treatment (be it diagnostic or therapeutic) is delivered and strong processes for careful measurement of key outcomes. Also, such trials frequently are conducted by persons with exceptional expertise who are following rigidly prescribed procedures. They are often referred to as *efficacy trials*. If a procedure cannot be shown at least to be acceptably safe and acceptably efficacious under such ideal circumstances, everything else about it is moot.

### Effectiveness

*As delivered, does it improve health?* The emphasis here is on the word “does”. The effectiveness of a procedure or process usually relates to whether its previously proven efficacy is actually realized in practical “field” situations and circumstances of use. Here, any or all of the restrictions or prescriptions applied in the efficacy trial may not apply. The subjects may be more diverse, the procedures less well applied and the outcomes less carefully measured. For these and many other reasons, it is common for such *effectiveness trials* to yield lower and more variable outcome performance than is found in the preceding efficacy trials.

### Equity

*Fairness of distribution and access.* Equity is a more tenuous and socio-culturally complex construct than ef-

---

ficacy or effectiveness. A somewhat simplistic equity-driven statement, for example, is that “... every subject in need should have equal access to a level of effectiveness that is the highest possible, given the resources available and each subject’s condition and attributes ...” In practice, what this means is that, to the greatest reasonable extent, variations across families in terms of *access* to services and *effectiveness* of services received should be minimized. The quality of care that a child and family receive should not depend upon which audiologist they happen to encounter or upon where they happen to live. This is a position that to many people is so obvious and fundamental as to be axiomatic.

### Efficiency

*Maximum improvement for minimum cost.* Efficiency seems relatively straightforward, at least at first sight. For a given quantity of resource, how can the maximum amount of benefit be obtained? The challenge comes in the attempt to quantify benefit. For example, in screening itself, an obvious measure of benefit is the achieved coverage of the newborn population. But even with this very simple concept, complications arise. At what age does screening become too late to qualify as “timely”? What expenditure of resources is appropriate to chase down the last 2% of newborns who may be very difficult to contact, may be located in remote areas or whose families may be equivocal about the merits of screening due to cultural or cognitive factors. In diagnostics, how much resource should be expended to track down families lost to follow-up? How complete must initial diagnostic results be in order to get on with, say, provision of appropriate amplification? In intervention, what levels of language performance, at what age, constitute “success”? The questions and issues in definition of benefits from program components or the program as a whole are almost endless.

### Legitimacy

*Conformity to social values.* In many ways, legitimacy is the most crucial facet of quality, in the sense that if society at large (the *people*, the *person in the street*, the “ordinary man” [sic], etc.) is not persuaded of the “rightness” of a particular healthcare program and the need for it, then the justification for it has to be questioned profoundly and the political will to fund it is unlikely to be forthcoming.

### Acceptability

*Conformity to personal values.* The personal values that really matter are those of the recipients of care. If any procedure, for example, is deemed to be too unpleasant, painful or difficult to tolerate, in relation to its perceived potential benefit, then by and large, it will not be utilized by those individuals for whom it is proposed. This aspect of “acceptability” has prominence in the widely-accepted WHO criteria for implementation of population screening for any disease or disorder (Wilson and Jungner 1968; Andermann, Blancquaert, Beauchamp and Dery 2008).

### Optimality

*Best balance of costs and benefits.* Costs may include direct and indirect costs of many kinds. So-called *opportunity costs* must also be considered; these costs relate to unavailability of resources spent on activity X to support an alternate activity Y, so that the loss of projected benefits of Y is in a sense a cost of X. Benefits include the entire spectrum of valued outcomes of the program. If those benefits can be expressed in monetary terms, then it may be possible to conduct a true cost-benefit analysis. However, it is notoriously difficult to monetarize some of the outcomes of EHDI, such as informed and empowered families, guilt avoidance, earlier access to sound or to alternative language modalities, specific receptive language scores at specific ages, changes in self-image and employment opportunities. If valid monetarization is not feasible, then resorting to cost-effectiveness analysis is necessary. Often, as is the case currently for EHDI programs, a blend of cost-effectiveness and cost-benefit analysis is necessary (see, for example, Grosse and Ross 2006).

### The “Three Es”

Of all of these facets of quality, three of them stand out as of special importance in optimization of healthcare programs. These are Effectiveness, Equity and Efficiency – what we might call the big three Es (“E3”). Some of the other attributes, such as Legitimacy and Acceptability, must be present in sufficient measure to justify program introduction. Cost-benefit balance is typically projected as part of the process of justifying EHDI programs to funding agencies. But, once the program is implemented, its actual value depends critically on E3. Also, within E3 there is a hierarchy. Effective-

---

ness is the crucial element. Without some reasonable level of effectiveness, a program will not survive and, indeed, it should not. The big question is: how can effectiveness be optimized, both at the level of the individual child and family, as well as at the level of the entire group of recipients of any services within the program? Equity is intrinsically a group concept, in that it reflects the distribution of effectiveness across the entire set of recipients of any specific service within the program. For example, does a 99% screening coverage for white, middle-class, urban families extend to non-white, economically disadvantaged, rural families? Do families in region A receive diagnostic services of similar timeliness to those in region B? Do the clients or patients of audiologists X and audiologist Y receive diagnostic audiometry of similar accuracy? The list of equity questions is large.

Equity is prone to interpretation in the light of the mores, values and priorities of any given society. In Canada, for example, it goes without saying that inequity of healthcare services is almost universally considered to be intrinsically unfair and deeply inappropriate. The right of any individual to choose to receive a basic level of health services that is unrelated to race, religion or socioeconomic level is a national axiom that is beyond questioning. In other developed societies, such as the US, viewpoints on equity are more diverse, and this is manifested in the realities of access to care. Whatever one's viewpoint, though, it is inescapable that if any subgroup does not achieve a desired criterion for success, there is clear scope for improvement in the *population* effectiveness of the program by increasing the level of equity. Thus, effectiveness and equity are inseparably intertwined at the population level.

For any given program service component and associated measure of effectiveness (such as the proportion of newborns screened, age at successful diagnosis, etc.), the effectiveness-equity relationship can be thought of as having two main characteristics: a general level of effectiveness, such as might be reflected in the population median level, plus the spread of effectiveness, such as might be reflected in the range between the 10<sup>th</sup> and 90<sup>th</sup> percentiles of the distribution of the measure. The smaller the range, the more uniform the distribution of effectiveness across individuals and the higher the equity. Another view would focus on the fraction of the population with the lowest effectiveness, arguing that while exceptional effectiveness is fine, an exceptional lack of effectiveness for some people in need is far more important.

## **Good Clinical Performance is not Good Enough**

EHDI programs are especially challenging because their purpose is not merely to screen babies but to deliver specific outcomes that may involve a sequence of several procedures, such as *screening – diagnostics – amplification – language interventions*. Each procedure has its own effectiveness aspects and measures that may be quite complex in themselves. For example, how would you create a measure or coefficient of diagnostic effectiveness? It would have to include aspects of diagnostic error (such as an incorrect threshold level), failure to make a key inference (such as type of hearing loss), and lack of timeliness (such as might be due to difficulty in getting complete test results). However, suppose these complexities could be overcome and effectiveness could be boiled down to a simple binary (yes or no) proportion, such as 90% of cases who required diagnostic assessment achieving diagnostic “success”. Suppose also that such an approach to effectiveness definition could be worked out for each of the four major procedural steps from screening to intervention. The next point, that the transfer or linkage of affected children and families from one procedure to the next, is every bit as important as the effectiveness of the clinical procedures themselves. For example, if 10% of infants who fail the screen never attend the initial diagnostic assessment, then if the likelihood of attendance is unrelated to true hearing status, that shortfall is equivalent to a 10% decrease in diagnostic success. The “net effectiveness” of the diagnostic stage is the product of the probability of attendance and the diagnostic effectiveness coefficient itself. This concept can be applied repeatedly to define “cumulative program effectiveness” up to and including every major step (procedure or linkage) needed to yield any specific, desired program outcome, starting from screening coverage onward. Moreover, if the screening itself has multiple steps, the concept of cumulative effectiveness can be applied to each stage as well as to the entire screening test sequence, in which case the probability multiplication can expand to five or seven elements up to entry into diagnostics! For more detail about this issue and many other aspects of a programmatic approach to EHDI, see Hyde (2011).

The point of putting this development in an article about protocols is simply to emphasize the pressure upon audiologists conducting any individual test component of the overall program sequence. Because EHDI is sequential, the effectiveness needed for individual pro-

---



cedures is extraordinarily high. For example, one might consider a diagnostic testing success rate of 90% to be "good" performance in the context of a conventional practice – of ten infants in the clinic waiting room, diagnostic success was achieved in nine of them. Similarly, screening staff could congratulate themselves on a 90% success rate for screening, and the program manager could congratulate herself on 90% newborn coverage and 90 successful follow-up of screening failures. The problem, of course, is that the cumulative effectiveness at diagnostic output is now 0.9 to the fourth power, or 0.66! That is, of all newborns with significant PHL, only two-thirds of them are getting a timely and accurate diagnostic assessment. Thus, it becomes painfully clear that from the perspective of the quality of the EHDI program as a whole, 90% effectiveness for any single step such as diagnostic assessment or provision of amplification is not nearly good enough.

## **Program Evaluability**

In most jurisdictions, it is accepted as a given that any major healthcare initiative will be subject to program evaluation. This will be familiar to most audiologists who work in a teaching hospital setting that is subjected to accreditation. Program evaluation is a clear component of Andermann and colleagues' (2008) update of the classic and widely-accepted WHO screening criteria originally developed by Wilson and Jungner (1968). Program Evaluation (PE) is a standard component of Quality Improvement (QI) processes, which are themselves widely considered a mandatory part of any program that seeks to demonstrate accountability and proof of quality.

PE and QI are dependent on the ability to quantify and evaluate procedures, to aggregate outcomes and identify potential sources of improvement in performance. It is impossible to quantify procedures that are not adequately described, to evaluate procedures that are highly variable in multiple aspects, or to summarize in a valid manner individuals' outcomes that are heterogeneous or that arise from unknown diagnostic or interventional processes. For example, it would be impossible to evaluate a process for diagnostic assessment or for provision of amplification, if that process were not well-defined or were not actually what happened in field practice by multiple providers, each with their own view and practice. One could treat the program's overall diagnostic or intervention process as a "black box" and evaluate outcomes as a whole, perhaps, but the measurement of outcome is only the beginning of PE/QI. The next step

is to want to improve program performance, and you cannot do that effectively if you don't know what was done, if what was done could change tomorrow, or if it could differ non-systematically from provider to provider, place to place or patient to patient.

Practice obscurity and practice variability are the mortal enemies of program evaluation and if you can't evaluate a program, you can't go about improving it systematically and rationally. No PE/QI, no accountability. No accountability, no continuation of funding. That is an increasingly common refrain in an area as costly and prominent as public healthcare. A partial remedy for obscurity and variation of practice is judicious adherence to a well-constructed protocol or guideline. Such recipes for practice are not a straightjacket, provided that they are constructed and utilized sensibly. Yes, they do limit freedom of action, but a really good protocol mainly limits the freedom to make egregious clinical errors or to at the very least waste time and money, which most would argue are human traits and freedoms of dubious value.

## **Evidence-Based Practice (EBP)**

The factors that drove the evidence-based practice (EBP) movement of the 1990s were highly relevant to the predicament of the EHDI audiologist just outlined. Pressure to achieve the highest possible levels of service cost-effectiveness arose in part as a consequence of policymakers' concerns over ever-increasing healthcare expenditures. The cost-benefit of EHDI programs, especially for universal screening versus targeted screening of high-risk newborns, remains a much-debated issue, even today. Second, the 1990s was a period of explosive growth of access to an ever-increasing mountain of scientific and clinical information, which had reached the point of being overwhelming to the busy clinician, whatever the field of practice. Third, there was clear evidence from several fields and many contexts that diagnostic and therapeutic procedures used to address a wide range of disorders could vary dramatically from place to place and from provider to provider, so-called "area" variations in practice. Not only that, but practice variations were linked strongly to geographic variations in resource expenditures for given health conditions, and the relationships between cost and quality of outcome were often found to be weak or nonexistent! In a nutshell, providers of care for a given health condition often behave very differently, some cost much more than others and those who cost the most do not necessarily get the best results.

The family-centered (“patient-centered”) care “movement” developed over a similar time frame from growing understanding that the more care recipients’ wishes, values, behaviors and engagement are taken into account in the selection and delivery of their care, the more likely a successful outcome, especially in the context of chronic disorders for which rapid, total “cure” was not available. This will be painfully familiar to EHDI audiologists dealing with issues of family denial of their child’s PHL, lack of engagement with intervention, non-compliance with recommendations and loss to follow-up. There was much more to the growth of EBP than this, of course. A version of evidence-based care was being practiced by many providers long before the EBP trend achieved public prominence. But there is something deeply problematic about major variation in the processes, effectiveness and costs of care for a specified health condition – particularly from the perspective of the individual or family who happens to receive ineffective or unnecessarily costly care simply because of who happened to provide it to them or where they happened to live. It is not difficult to see the relevance of all this to basic questions of effectiveness, equity and efficiency.

The response of the health services community was a major movement in the direction of assembly, evaluation and distillation of “evidence” related to any specific area of healthcare. The deliverables were various kinds of “guidance” for clinicians. These took various forms and spawned all manner of terminology, including protocols, guidelines, practice parameters, practice standards and the like. All these are essentially recipes of one kind or another that differ in their style, scope, strength of compulsion, content and level of detail. Of course, they also differ in their own quality, a fact that was recognized early on and that resulted in guidelines and evaluation tools for guidelines, protocols for protocol development and, of course, a host of approaches to the evaluation and integration of primary scientific and clinical reports.

### **Was EBP the Answer to Concerns about Quality?**

Many would argue that the evidence-based practice movement has changed irreversibly the landscape of healthcare service provision. It has encouraged an ethos of critical evaluation of how best to achieve specific healthcare objectives and how to measure not just whether procedures were delivered, but whether they actually yielded the desired health outcome. It has given us dimensions and criteria for the quality of primary reports,

ranking scales for the strength of evidence, algorithms and conceptual frameworks for the process of evaluation itself, an entire industry devoted to the creation and evaluation of quality of care. Yet, despite the best of intentions and some singular successes in a few areas, by and large the EBP movement has had limited success in changing clinical practices. The big question is: why? First we will take a quick look at the major elements of EBP.

The methodology of EBP itself is largely and very deliberately transparent and self-critical. If it is seen as a process by which high-quality practices are to be discovered or developed, described clearly and delivered successfully into the outstretched arms of care recipients, then it can be said that issues and problems have come to light at every stage in that process. The process starts with published reports. For many reasons, most primary clinical reports are not useful. Even a cursory look at a few results of high-quality “systematic reviews”, whether in areas of audiology, speech-language pathology or elsewhere, should convince the reader that despite the best efforts of editorial staff, most published reports in clinical journals are likely to have significant methodological, analytical or reporting flaws. Unfortunately, ordinary peer review procedures do not address this quality problem effectively, and nor should they be expected to. Peer review is usually done *pro bono* in “spare time”, and by and large, in this world you get what you pay for. Reviewers are selected according to subjective and variable criteria, review protocols often do not satisfy reasonable EBP methodological standards, and it seems that the general level of critical appraisal training for clinicians and reviewers alike leaves much to be desired. How else can you explain the fact that so many publications fail to describe their study subject sampling frame, fail to calculate hypothesis-testing power and sample size requirements, do not discuss threats to validity and generalizability, present uninformative or frankly misguided statistical analyses (especially the perennial “regression 101” errors such as the “starry cluster and distant dot” error) and draw inferences that are unsupported by the data, meanwhile all of it cruises happily through peer review? The bottom line is that a published report that is NOT invalid is a rare and beautiful thing to be cherished. The challenge is: how to recognize such a thing?

The next level of challenge is the quality of efforts at collection, evaluation and synthesis of study findings. Historically, reviews were written mostly by aging sages generally deemed to be “experts in the field”. These opinion pieces range from inspired and penetrating through to utterly useless or seriously misleading. Again, the problem

---

is to tell the one from the other. Perhaps the acid test is to take one or two review statements or recommendations by any given so-called expert, put them to the test in daily practice and if they don’t work, email the expert asking for an explanation. Persistent lack of response, gobbledygook response or a peremptory “this is how I (we) do it” should speak volumes about credibility.

Fortunately, EBP has provided sensible guidelines and protocols for the proper conduct of “systematic” evidence reviews. The interested reader is directed to websites such as that of the Cochrane Collaboration ([www.cochrane.org](http://www.cochrane.org)), the Methods Guide of the US Agency for Health Research and Quality (Google “AHRQ Methods Guide”) and the Center for Evidence-Based Medicine ([www.CEBM.net](http://www.CEBM.net)) in Oxford, UK. One major methodological missile in the EBP armamentarium is meta-analysis, a process by which the findings of many studies can be assembled and analyzed as an integrated whole. When done correctly, meta-analysis can be immensely revealing, both about sources of bias or of conflict in primary studies and about the nature of underlying truth that may not be apparent from any of the individual studies considered, mainly a consequence of insufficient sample sizes. Two of the major issues in meta-analysis relate to the way in which candidate studies are selected for inclusion and the way in which quantitative data from individual studies are combined. A central challenge in meta-analysis is the manner in which inferences based on aggregated data lose their accuracy and generalizability when the contributing studies or actions become less and less homogeneous. How valuable, for example, would be an aggregate analysis of ABR or ASSR threshold properties obtained under differing stimulus conditions or in differing populations of subjects? This relates back to the points about difficulty of evaluating activities that are poorly defined or highly variable. How valuable would the vast global pool of EHDI program outcome data be if all procedures in all programs were well-described, evidence-based and were substantially similar in many respects, permitting meaningful aggregation of data and evaluation of systematic differences in procedure? The interested reader is referred to introductory texts such as that by Lipsey and Wilson (2001). For much more detailed reviews of EBP and its impact in pediatric audiology, see Gravel (2005) and Hyde (2005).

## Knowledge Transfer

Now we turn to protocols and guidelines themselves. In this chapter, for simplicity both clinical proto-

cols and clinical practice guidelines will be referred to as “protocols”, acknowledging that in conventional terms protocols typically are more specific, include more procedural detail and may have a more obligatory nature. There are a spectrum of these attributes, grey zones and no generally accepted, clear dividing line. As might be expected, the field of EBP provides many sources of guidelines on how to produce well-designed practice guidelines and how to evaluate such guidelines. The reader is recommended to Google the “AGREE Collaboration” for further information.

At this point, suppose that we have a well-written protocol based on solid systematic or semi-systematic reviews of primary reports, with appropriate selection and grading of reports and synthesis of evidence. The reader might think that one is a hair’s breadth away from improved clinical practice, but that is not so. *Production* of a high-quality protocol is just the beginning. The challenge is to transfer the knowledge reflected in the protocol into the thinking and daily clinical practice of healthcare providers. It turns out that this process is every bit as complex as the production of a good protocol in itself. One reason is that knowledge tends to be based on observed “facts”, whereas clinical practice reflects beliefs, attitudes, aptitudes and operational contexts.

The reader will not be surprised that studies of “guideline utilization” and of “barriers to guideline use” are hot themes currently in the EBP world. The buzzwords include “knowledge transfer or translation (KT)” and “change management”. One of the most influential reports to date on the topic was by Cabana et al. (1999) on barriers to “adherence” to clinical practice guidelines. The main barriers they identified are listed below, along with the present author’s re-ordering and translation of Cabana et al.’s terminology.

1. Lack of awareness – *Guideline? What guideline?*
2. Lack of familiarity – *Haven’t read it. Impenetrable.*
3. Previous-practice inertia – *I know what I’m doing.*
4. Low outcome expectancy – *No point. It won’t work.*
5. Lack of agreement – *Wrong, irrelevant, cookbook.*
6. Guideline barriers – *Hard to use, inconvenient.*
7. External barriers – *Haven’t got the time.*
8. Patient barriers – *Patients don’t like it.*
9. Environment barriers – *Not enough resources.*
10. Lack of self-efficacy – *I can’t seem to manage it.*



The list is diverse but grounded in evidence, and there is logic to the sequence. Complete lack of awareness of any guideline (1) renders the question of its utilization moot. It could be argued that the health professional has an obligation to be reasonably current in awareness of guidelines. Next (2), a provider may be aware of a guideline, may have decided to ignore it, have intended to review it but not had the time, or have tried to read it and found it difficult to understand. Item (3) reveals a level of arrogance that is difficult to justify and that, in the author's experience, is often associated with limitations of practice quality. The true professional is likely to be well aware that hubris is for teenagers, that it's never too late to try something new, that standard practices may have ossified over time or been flawed all along and that the day you stop clinical learning is the day you should consider a second career.

Items (4) and (5) reflect the result of some degree of personal evaluation of the guideline, which may range from cursory dismissal after a quick read to a genuine attempt to use it, with poor results. In (5), "wrong" may reflect a perceived flaw of concept, "irrelevant" may be a snap judgment or may reflect a genuine mismatch between the provider's population and the populations on which the guideline was based, and "cookbook" may reflect either inappropriate non-differentiation of patient/client groups within the guideline or a provider blanket belief that healthcare is just too complex and individual patients just too idiosyncratic for their care to be well-specified by categorical rules. Such an uncritical position may have more to do with self-image than actual evidence of pandemic practice complexity.

Items (6) to (9) represent an array of barriers each of which has the potential to be entirely valid in a specific context. Alternatively, any one may reflect a biased viewpoint, and an issue that may well arise is that of practice effects: it is easy to gain a false impression of the utility of a novel approach, simply because it may take time and experience to do it properly. Another possibility is that belief (or lack of it) in a positive outcome may actually affect the likelihood of such an outcome.

The interested reader is referred to Francke, Smit, de Veer and Mistiaen (2008) for a recent review on barriers and to Armstrong, Waters, Roberts, Oliver and Popay (2006), Straus, Tetroe and Graham (2009) and Ward, House and Hamer (2009) for broad coverage of the field of knowledge transfer in the healthcare context.

## **Lessons from the Ontario Infant Hearing Program (IHP)**

As a simplistic summary, the list of barriers to utilization can be thought of as reflecting three basic domains: the attributes of the protocol itself, attitudes and beliefs of the healthcare provider and the context of practice. At the time of planning (in 2000) for possible initiation of an EHDI program in Ontario, Canada, an attempt was made to put in place a system that would address at least some of the potential barriers raised by Cabana et al. (1999). Many lessons have been learned from that effort as it has played out over the last decade, with partial but by no means complete success.

The Ontario Infant Hearing Program (IHP) started service delivery in 2001, after about 20 years of previous screening and follow-up services directed at high-risk newborns. It is funded by the Ontario government as part of a suite of programs directed at enhancing early child development (Early Years' Programs). The suite includes a major program for Preschool Speech and Language, with which the IHP is strongly linked and with which it shares substantial information systems and administrative infrastructure. The IHP has centralized control of funding and policy but with regional administration. The funding is vertical, meaning that it addresses all stages of EHDI from initial universal screening through to longitudinal interventions to maximize development of language. The program does not fund directly medical actions, which are themselves funded by a universal healthcare system that adheres to general principles of universality that are mandated federally and largely funded provincially. Also, the capital costs of hearing aids for children are defrayed by a separate government program, the Assistive Devices program.

Ontario has an area of 400,000 square miles but a population of only about 13.5 million and an annual birthrate of about 132,000. The IHP has gross performance statistics typical of a "good" EHDI program: coverage in the mid-90s percent, refer rate to diagnostics 1.5%, average yield of congenital PHL around 2/1000, follow-up rate to diagnostics in the mid-80s percent and so on. What is most unusual about the IHP is its effort at optimizing practice quality and program-wide consistency. For example, all audiologic assessments and amplification provisions are done with identical instrumentation, test parameters and operating supplies. All audiologists who provide IHP services must be registered with the Ontario College of Audiologists, and Speech-Language Pathologists must receive post-certification training and

---



performance monitoring by the IHP and must follow detailed, evidence-based protocols that are reinforced by clinical decision support from IHP centers of excellence and various kinds of performance audit. About 85 audiologists currently provide clinical IHP services. This makes the IHP a natural experiment for study of the pursuit of quality of clinical services in a protocol-driven context.

The IHP’s current protocol for initial diagnostic assessment in infants who fail newborn screening is available or has been in place (with minor modifications) since 2001. It includes diagnostic Distortion Product Otoacoustic Emissions (DPOAE) testing (nominal f2: 1, 2, 3, 4 kHz), diagnostic tonepip Auditory Brainstem Response (ABR) testing by air conduction (0.5, 1, 2 and 4 kHz), bone conduction (0.5 and 2 kHz) ABR where indicated (with two-channel recording), middle ear analysis (MEA) including tympanometry with a 1 kHz probe under 6 months and 226 Hz over 6 months and ipsilateral middle-ear muscle reflex (MEMR) recording (1 kHz stimulus and probe). Auditory Neuropathy Spectrum Disorder (ANSD) is assessed using rarefaction/condensation click ABR, cochlear microphonics (CM) and summing potentials (SP), as well as DPOAE.

Some decisions about which tests or variations should be included in a high-quality diagnostic assessment protocol are straightforward, but not all. For example, on reflection it is clear that click ABR testing alone is not a sufficient basis for audiometric description or prescription of amplification. Basing an infant’s hearing assessment on a click ABR, while much better than doing nothing at all, is comparable to conducting an adult hearing test using only a white noise stimulus, which most audiologists would consider to be absurd. Other decisions, such as those relating to choice between tonepip ABR and ASSR or to their complementarity in an integrated protocol, can be much more complex. Often, the devil is in the details. For example, tonepip ABR threshold estimation varies enormously in accuracy from tester to tester, depending on many details of stimulation and recording parameters, test environment, baby management, test strategy and tactics, interpretive skill and experience and caseload volume and diversity.

The IHP protocols for diagnostic audiologic assessment and for provision of amplification are unusually detailed. For example, the former is over 80 pages in length. The protocol is based on many semi-systematic evidence reviews, which are routinely funded by the IHP as part of its QI effort. A semi-systematic review includes explicit document database search and citation inclusion criteria, but may lack the formality of a full, sys-

tematic review, for example in relation to multiple-reviewer citation prioritization and conflict resolution. A major advantage of achieving at least the semi-systematic level of review is that the process is defined and reproducible, unlike reviews based on opinion-driven or covert inclusion criteria.

IHP protocols include both mandatory and discretionary components. Mandatory components are those elements that IHP has determined **MUST** be done, either on the basis of evidence review or, where evidence of an adequate standard is lacking, on the basis of expert consensus. The latter are kept to an absolute minimum because of their intrinsically subjective nature, but are necessary because good theoretical justification or experimental evidence is lacking for so many important, even fundamental, aspects of test protocol and interpretation.

IHP audiologists are able to contribute to protocol development and evolution, but the process for such change is systematic and the resultant modification is applicable to all IHP audiologists province-wide. If any protocol element is challenged or any error or omission identified, the issue is reviewed by IHP consultants as well as external experts in the field. New evidence review may be deemed necessary and IHP audiologists may be surveyed on any issue. The audiologist who raised the matter is then provided with a response, an explanation of why the matter will not cause protocol change or that it will do so and in what way.

## IHP Protocol Characteristics

IHP protocols do not simply tell our audiologists what to do. There are so many variables at play in service delivery, be it diagnostic or interventional, that it is often not possible to deal in depth with each and every circumstance that may arise. Nor is it desirable to do so, because it will render the protocol hopelessly unwieldy and impenetrable, as well as tending to diminish independent thought and exercise of judgement by the audiologist. A judicious, intermediate path must be navigated. One tool for such navigation is to emphasize not the minutiae of actions but rather the *principles* that should guide those actions. To illustrate this, a few of the key concepts in the IHP protocol for diagnostic assessment, for example, are outline below.

### *Limiting Test Objectives*

Proof of the presence or absence of a program “target” PHL is critical. Reflecting the well-known WHO

---

screening criteria (Wilson and Jungner, 1968), target PHLs typically should not include those PHLs that cannot be detected reliably, those for which evidence of impact (i.e., does it matter?) is lacking, and those for which no effective intervention exists. Typically, EHDI program target PHLs have lower limits of 30–40 dB HL in the range 0.5 to 4 kHz. In the IHP, measurement of hearing sensitivity below a lower limit equivalent to 30 dB HL is irrelevant and is not funded. It is important to grasp the programmatic principle that scarce resources are *not* expended in activities that are not necessary by program objectives definition. Not only that, but measuring thresholds below 25–30 dB with ABR or ASSR raises a host of difficulties and departs from any hope of effective and efficient practice.

### *Optimizing Information Gain Rate*

ABR-based (or ASSR-based) test strategy in young infants is (or at least should be) very different from standard protocols (ANSI or ISO) for audiometry in adults, in which it is presumed that a defined, “complete” set of thresholds will be obtainable. In infant diagnostic testing the period over which EEG noise levels are satisfactorily low may be less than desired and its end may be unpredictable. This is always the case when testing in natural sleep, but even when common sedatives are used the limitation and unpredictability are not always eliminated. Second, there is no guarantee that the family will return for subsequent appointments. Loss to follow-up is a well-known challenge and many factors may limit a family’s intent or ability to attend repeatedly. This adds up to a need to maximize the rate of audiometric information gain.

*The key principle is always to behave as if testing may be terminated in the next minute.* This leads to a “top down” strategy of progressive refinement of diagnostic information. The question is always: *given what I know right now, what next test condition will give the highest expected gain in diagnostic information?* There are two aspects of the answer: the information content, such as threshold normality or loss severity, frequency range, site of lesion, etc., as well as the probabilities of the possible outcomes. For example, if only 10% of the test population were likely to have PHL, it makes little sense efficiency-wise to start ABR testing at a moderate intensity level. Rather, it should start at whatever level immediately identifies normal hearing. A more difficult strategic question is whether diagnostic testing should always start with bone conduction ABR, because immediate

proof of absence of sensory PHL is highly informative. The counterargument is that proof of absence of any hearing loss by air conduction testing is even more informative, so start with AC testing. But, what if transient, conductive loss were very common? Then, many children would be abnormal on AC and so BC would be necessary to rule out a sensory component, and so on.

Another key issue is the ABR tonepip frequencies and their order of testing. Is 2 kHz the single most important test frequency? If 2 kHz testing were successful, then what is next – 4 kHz or 0.5 kHz? Is 1 kHz necessary if 0.5 and 4 kHz give equal thresholds, etc? Similarly, if an infant shows no ABR at initial testing at the minimum required intensity for any given stimulus frequency, raising the intensity by 10 dB steps is highly inefficient unless mild hearing loss is far more likely than greater degree of loss. In general, a more efficient strategy is to ascend in steps that are much larger initially (such as in one or two steps of 30 dB), to reach the threshold region as quickly as possible, before bracketing the threshold with smaller steps. There are many such questions of optimal test efficiency. Cumulatively, failure to optimize can cause major cost overruns, increase loss to follow-up and compromise timeliness of diagnostic completion and initiation of interventions.

### *Do NOT Guess! Making Rational ABR Detection Decisions*

From the author’s experience providing clinical decision support and conducting ABR quality audits over the last decade, there is a strong tendency for audiologists to treat the judgment of ABR presence or absence in any given average as a forced-choice detection problem. It is as if the ABR were either there or not there, yes or no, and once the binary judgement is made, it is believed to be correct. This approach can lead to an unconscious tendency to make guesses about ABR presence or absence in averaged records. The true situation is that each response judgment is a statistical hypothesis test. The “null” hypothesis is usually that response is absent. To decide that a response is present, on statistical grounds it is usually sufficient that the size of the putative “response” be at least about three times that of the standard deviation of the EEG noise level in the averaged record. An issue is that the noise levels can vary massively across subjects and across averages within subjects. To a first approximation, averaging with a fixed number sweeps (N) simply reduces the standard deviation of the averaged EEG by a factor of root-N relative to

---

the raw EEG. It follows that a five-fold variation, say, in EEG noise levels across subjects translates directly to a five-fold variation in averaged noise levels. Therefore, using a fixed number of sweeps and a criterion for ABR presence of three times the averaged noise level will result in large variation across subjects in the absolute ABR size that will trigger a judgment of response presence. This is obviously inappropriate and will contribute to variation in accuracy of ABR threshold estimates.

High-noise EEG is mainly caused by electromyogenic interference, and the first-line remedy is to improve the baby’s state. The second line is to ensure that electrode impedances are as equal as possible and are reasonably low. The third line is to ensure appropriate recording bandwidth and artifact rejection levels. Given all of that, the final resource is to vary the size of averages in order to promote more uniform final noise levels in the averaged EEG. Very small averages should be avoided because their estimated noise levels are quite unstable. Very large averages should be avoided because they are inefficient in term of gain in noise reduction per unit test time, because of the diminishing returns that follow immediately from the root-N law. In the author’s opinion, a range from 500 to 4000 sweeps is reasonable and will allow almost a three-fold variation in EEG noise levels to be accommodated so as to produce a consistent averaged noise level. This means that if the average rapidly develops a response-like waveform and that waveform is at least three times the size of any other deflection, then a response-present decision is reasonable, regardless of the absolute EEG noise level and even if only 500 sweeps have accumulated. If a smaller candidate response is seen, then the absolute noise level *does* matter, and averaging should proceed to a reasonable upper limit, to try and lower the averaged noise to some acceptable criterion value. A suggestion is that the criterion value should be that which is typically obtained, under the exact recording conditions used, for averages of about 2000 sweeps in babies with “very good” EEG.

To make reliable decisions that response is *absent* requires even more attention to noise levels than making decisions that response is *present*. *Here, the key principle is that the decision of response absence is valid only if conditions are such that a genuine response would have been detected with high probability.* This means that a maximum permissible averaged noise level criterion must be applied to all averages for which a response-negative decision is contemplated. One question is: what happens if neither the “response present” criterion for

response signal-to-noise ratio nor the “response absent” criterion for absolute averaged noise is met within a reasonable maximum recording time? This can happen frequently. The point is that the response detection decision is *not* binary. The third option is essentially “cannot decide” or “indeterminate”. It is far better clinically to assign an “indeterminate” outcome to a given stimulus condition than it is to force a judgement of response presence or absence when the recordings simply do not justify such certainty.

An underlying issue is how to estimate the averaged EEG noise level quantitatively. In the past, testers were limited to a subjective impression of the size of the fluctuations in the average in latency regions in which response is not expected. More recently, equipment manufacturers have incorporated a calculation and display of the noise level in the average as it accumulates. The calculations differ among manufacturers, unfortunately, but they do at least provide a device-specific numerical noise estimate. Such estimates are commonly referred to as “residual noise levels” or RNLs. In the author’s view some kind of RNL computation is at present an essential aid to reliable, subjective response detection judgments, in the absence of proven valid and powerful automated response detection algorithms that are appropriate for time-domain evoked potential waveforms in the context of diagnostic testing.

These are only three examples of the many aspects of the IHP diagnostic protocol that its developers considered to be essential for promotion of the three key facets of service quality (effectiveness, equity and efficiency) in relation to ABR-based diagnostic assessment. Other important areas include the determination of the responding cochlea in BC tonepip ABR, methods for dealing with electromagnetic power line artifact and stimulus artifact, rational strategies for evaluation of ANSD, etc. One important point is that telling audiologists that they must use, for example, tonepip ABR to estimate thresholds in babies who fail newborn hearing screening is a small step that is just the beginning of a useful protocol. The real challenge and the real utility come in the explanation of why that approach is appropriate and how to do it in a way that is valid, practicable, accurate and efficient.

## Lessons from IHP Protocol Training

Audiologists who are authorized to be trained to conduct IHP diagnostic assessments receive an individualized, three-day hands-on training program. The one-on-

one diagnostic training on the IHP equipment and protocol starts with a three-day course that includes a review of prior training and experience, a course needs assessment, protocol orientation and clinical sessions with six babies. An attempt is made to schedule cases in such a way as to provide a progressively complex and diverse hands-on experience. Early experiences attempting to train more than one audiologist at a time were unfavourable, mainly because of differences among trainees in attitude and knowledge. Also, a “train the trainer” model was rejected out of hand, because of a viewpoint that such models might be vulnerable to progressive distortion of key messages, such as can occur in the party “whisper game”.

As mentioned earlier, on starting IHP service provision after the hands-on training phase, the audiologist trainee is required to submit all clinical records for rapid review of technique and interpretation by a center of excellence, prior to issuing clinical reports. This review phase of training lasts until each audiologist’s testing and interpretation are considered to meet a required minimum standard. Thereafter, audiologists may discretionally seek expert review of any records they wish, such as for unusual or difficult cases. In addition to the post-training monitoring, any IHP audiologist may be required to submit records for mandatory audit, either by random selection or driven by any adverse event that comes to light, such as a complaint by a family or an unresolved disagreement among IHP audiologists involved in any given case. Such a highly regulated context of practice may seem extraordinary and even unacceptably intrusive to many who are not familiar with the accountability demands of programmatic services. It is to the great credit of the audiologists of Ontario that on the whole they have shown an understanding and acceptance of the need for such a system. The performance monitoring and audit processes are managed in a manner that is confidential and intended to be supportive rather than threatening or punitive. However, the crucial point underlying all of it is that it is the interests of the child and family that are paramount, not the ego of the provider or the prestige of a specific profession. In any case, it can be argued that there is little that benefits the status and satisfaction of a health professional more effectively than to be, and be seen to be, the agent of the highest possible quality of care. And as the program flourishes, so may its providers.

In connection with training, one important challenge that has arisen in Ontario is that the pendulum of access to services has swung too far from isolated centers of excellence towards local access everywhere.

There is a compromise to be struck among access, caseload and skills. Over 80 audiologists currently provide IHP ABR-based diagnostics, but fewer than 2,500 initial diagnostic assessments are required annually province-wide. In that 2,500 there will be only about 150–250 cases of congenital PHL and these will include only about 20 cases of ANSD, for example. If caseloads were distributed uniformly across all trained audiologists, each provider would do only about 25 initial diagnostic tests per year, any given audiologist might see only a handful of genuine PHL cases or difficult cases per year and an ANSD case once every few years. Such sparse caseloads pose an inherent challenge of skills development and maintenance, even with detailed protocols and supports. This situation is one reason why training of new IHP audiologists is now restricted to situations of absolute geographical necessity. It is also a justification for a tiered system of diagnostic referral to regional centers, as well as the imposition of minimum caseload requirements for diagnostic and intervention service providers.

While not designed as a formal experiment, these experiences of training and performance review have provided many insights. Note that the observations below are based on subjective inferences from our specific experiences and may not generalize to all training and service contexts.

### *Experience and Resistance to Change*

In general, we have found that the less practice experience audiologists have in infant ABR testing, the more receptive they will tend to be to new ideas and approaches. Some very experienced audiologists accept new ideas, techniques and protocol requirements readily and rapidly become testers of the highest quality. In contrast, audiologists who are resistant to protocol obligations tend to have many years of experience in classical pediatric audiology but not, inevitably, in newborn, frequency-specific ABR with AC and BC stimuli.

### *Individuality of Training Needs*

Despite commonalities in their academic training, audiologists who have attended IHP training are highly diverse with respect to their level of understanding of physiological measures and related techniques, their training needs in terms of content and style of learning and their response to different approaches. They are also diverse in their approach to response detection judgments, especially in relation to their conservatism

---



and decision speed, probably a corollary of personality variables. They also have idiosyncratic aptitudes and conceptual "blocks". Overall, the lesson is that training in advanced diagnostic techniques appears to need to be highly individualized, which is not surprising.

### *Commonality of Conceptual Challenges*

Despite the diversity of individuals, certain concepts tend to need careful thought and explanation, re-explanation and illustration with practical examples. The avoidance of inappropriate threshold-seeking, the principle of optimizing information gain and the importance of not guessing about response presence or absence are three areas that appear to need much explanation and re-enforcement.

### *Hypnotic Averaging Syndrome*

There has been a common tendency for trainees to be insufficiently engaged in the process of averaging. This is not surprising, given that most of our trainees had more experience with otoneurologic ABR applications in adults, for whom the process of averaging is akin to watching television and active engagement is not necessary. Threshold measurement is quite different in its complexity and conduct. For example, in babies who do not sleep soundly for an entire session, the overall quality of averaged ABR records depends on appropriate settings of artifact reject levels. It is typical to conduct averaging with artifact reject limits that are far too wide, given that an orderly average perhaps showing a clear response can be disrupted irreversibly by only a few sweeps of activity with high myogenic artifact. Even if reject levels are sensitive enough that most large artifacts lead to sweep rejection, there are typically build-up and tail-off phases to myogenic artifact bursts in infant EEGs. This results in the inclusion in averages of sweeps that have noise levels that are high but that do not quite meet artifact rejection criteria. Much of this problem can be avoided by close attention to the baby and to the ongoing EEG, with interruption (pausing) of averages as soon as a myogenic burst appears or is pre-saged by overt movement, and with resumption of averaging after a return to quiet EEG.

### *An Obsession with Inefficiency*

One area that tended to occupy much of the case monitoring effort was the optimal choice of stimulus

levels for rapid threshold bracketing. It seems that many audiologists have been trained or have become indoctrinated to employ small intensity steps and to develop input-output functions. This type of recording and display is commonly displayed on websites and equipment manuals but is hopelessly inefficient for threshold estimation. Moreover, IHP protocol takes very little account of latency input-output functions, because such functions are of little or no diagnostic value in a context of tonepip ABR with both AC and BC measurements.

### *Charts, Charts, Charts...*

One activity that is widely appreciated by trainees is chart review. Several hours are now devoted to choosing infant ABR charts at random and asking the trainee to describe what they see, what it means and why, whether a procedural or interpretive error was made, and so on. A frequent comment is that such diverse, interactive decision-making under pressure is far more informative and realistic than even the best didactic teaching.

### *Initial Performance Monitoring is Essential*

One cannot learn to be a skilled diagnostician in three days, one reason being that the range of situations encountered is vastly greater than that which can be constructed in hands-on training. Very few audiologists managed to achieve optimal test strategy and tactics immediately, which is not surprising given the many factors involved.

### *Lessons from Elective Decision Support and Obligatory Audits*

Once the audiologists have completed the IHP hands-on training and case review phases, they are able to test and report as they see fit, but they are encouraged to seek a review opinion if they are uncertain about a given pattern of findings, about what to do next, or if they feel that a case is not well-served by the IHP protocol. Audiologists also are contracted to comply with a random audit process of program quality improvement. In that process, the auditee chooses some (usually four) cases and the center of excellence chooses another four cases, and all records and reports of the eight cases are reviewed in detail for errors and omissions and for protocol adherence. In addition, there is a formal process for "serious adverse events", which are events that come to light by a variety of mechanisms and which appear to

---

have significantly compromised clinical services to an individual case or group of cases. The main subjective inferences from these processes are that most audiologists do adhere to most key elements of the IHP protocols, but that diagnostic efficiency is an enduring challenge, with inefficient intensity sequences and averaging being commonplace.

Only half of all IHP audiologists use clinical decision support. The most frequent users tend to be audiologists whose work is already of very high quality. *The concern, of course, is the testing quality for those who never make use of decision support.* Random audits to date have revealed occasional, radical deficiencies of test practice and of clinical inference.

There are likely to be individual audiologists who consistently do not adhere to IHP protocol. The challenge is how to identify them and remedy the situation. Respect, debate, evidence and persistence are likely to be useful ingredients of an effective approach. In some contexts, programs are structured in such a way that the ability of an audiologist to provide program services can be revoked if reason fails to prevail, but such “sticks” are always a last resort. The justification for enduring refusal to adhere to a protocol is puzzling, given that there are clear and unused mechanisms to modify the protocol to the benefit of all concerned. The remedy for such situations is work in progress. Over the last decade, a few audiologists have withdrawn from IHP practice in a context of quality concerns.

*A handful of serious adverse event audits to date have revealed in every case a major violation of IHP protocol.* An egregious case, for example, would be a child with recurrent middle ear disease and conductive loss being fitted with hearing aids without bone conduction ABR ever having been done to confirm a sensory component.

## **Bottom Lines – Attributes of Good Protocols**

### **Address in Detail where the Rubber Hits the Road**

The world of EHDI is replete with protocols and guidelines. In the US, for example, almost every state has a protocol for diagnostic assessment. There are remarkable differences among these protocols, though many of them are based to varying degrees on the Joint Committee on Infant Hearing’s various guidelines (see, for example, JCIH 2007). What is more remarkable is that many of these protocols are only one or two pages

long. Some contain statements such as “Thresholds should be measured by a frequency-specific technique such as tonepip ABR or ASSR”, or “Diagnosis of ANSD should include recording of cochlear microphonics to rarefaction and condensation stimuli.” Such statements are necessary but are *completely* inadequate to help the audiologist do a good job. Nor can we fall back on the argument that audiologists learn to use these techniques in graduate programs. How many graduate programs in the US, for example, provide in-depth knowledge and practical skill in frequency-specific ABR testing? And even if they do teach it, how many audiologists graduate having tested fewer than five babies? And how many babies does it take before an audiologist can be said to be highly skilled and experienced?

### **Include Strong and Detailed Rationale and Justifications**

A good protocol portrays respect for the potential user. “Do it because I say so” might work for a time with young children but certainly will not work with highly-trained health professionals. It’s not about “do this”, it’s about “if you do this ... then this will happen ... and here’s why”. Even when something novel is explained well, most people will not accept it at face value. The challenge, therefore, is to convince audiologists that a particular procedure is worth trying out seriously, that it merits more than cursory dismissal, and that it actually might just be better than the old way.

The elements of rationale and justification may lie in published experimental or clinical evidence, in appeal to basic insight driven, say, by the laws of physics or statistics or, in the absence of anything more objective, may lie in expert opinion. But even expert opinion cannot stop at a bald statement of what should be done – it must include some attempt to explain why. It is not appropriate, for example, to justify a particular procedure by the statement “it works for us, so you should do it”. True expertise, in most cases, is not merely phenomenological but is based on genuine insight, and the challenge is to recognize the insight and be able to express it clearly.

There is, however, one area of considerable difficulty in developing protocols for programs. It relates to variations in audiologists’ aptitude, as well as a trade-off between equity and effectiveness. The use of protocols is at a minimum an attempt to avoid serious adverse events. One level up from that is an attempt to assure a minimum standard of care across all clients and all providers. The problem is that in focusing on ensuring

---

such equitable consistency, protocols may oblige providers to follow procedures that the inspired diagnostician, for example, might find tedious and unnecessary. An example of this is replication of averages in the course of threshold search. There are very good statistical reasons why judgments that an ABR is present should always be based on replicated averages. The argument has to do with the fact that in practice, EEGs are almost never well-behaved, simple random processes and that response-like waveforms can occur by chance with unexpected frequency. In that scenario, two averages of, say, 2000 sweeps, together with a requirement that the suspected response waveform should be present in both averages, are less likely to give a false-positive response detection than a single average of 4000 sweeps. If the EEG is truly well-behaved, splitting of averages is actually slightly less efficient than a single average, but in that situation it is usually no problem to achieve satisfactory signal-to-noise ratio in a short time. It is when the EEG is less than ideal that replication becomes an important tool.

Testers with great aptitude or great experience can recognize adverse EEG conditions and questionable responses to a degree that is difficult to explain or emulate with formal statistical decision rules. It is an art only part of which can be taught. Such testers could almost certainly achieve higher levels of accuracy and efficiency acting according to their own inherent skills than is possible using a protocol designed to guard against the effects of inexperience and misinterpretation. Overall, excellence has to be traded against variation in quality, in a context in which excellence is difficult to identify and quality variation is the norm. This is a balancing act that can be challenging.

### Where Decision Options Ramify, Stress Key Principles

There are dozens of key facets of doing a good tonepip ABR threshold determination, for example. Several of them could each occupy an entire three-day conference: stimulus parameters, recording parameters, averaging tactics, frequency strategy, stimulus route strategy, intensity strategy, response recognition, artifact recognition, overall clinical inference, and so on. Review of the literature will reveal a remarkable range of opinion and only a few sources of solid evidence. These are not hard to identify. But some aspects of test conduct are too complicated to describe fully. In such cases, it is very important that the protocol address core underlying prin-

ciples of test tactics and strategy. It is simple to justify selection of tonepip ABR, for example, but even more important to explain how to achieve high response detection decision accuracy and optimal test efficiency, both of which are crucial for program effectiveness, equity and efficiency.

### Minimize Mandatory Protocol Elements

Good protocols should distinguish clearly among procedural elements that are obligatory, those that are recommended but not obligatory and those areas for which there is no clear evidence or strongly-held program position. The general idea is that if you are going to ask people to do something with which they are not entirely familiar, insist on as little as is absolutely necessary. In the IHP protocol, for example, it is mandatory to do BC ABR at 2 kHz if the AC ABR threshold at 2 kHz is greater than 30 dB nHL. It is recommended that when the ABR threshold exceeds 70 dB nHL, it should be resolved to within 5 dB, whereas below 70 dB, 10 dB steps are sufficient. No position at all is currently taken about whether the test session should begin with OAE or ABR measurements, because many variables determine the best course in the individual baby, and the audiologist is the best judge of what to do. Similarly, for example, in the IHP amplification protocol, use of RECDs in prescription verification is mandatory, the routine determination of RECDs in threshold testing is recommended because of longitudinal, developmental effects on effective stimulus SPL at the tympanic membrane for a given dB nHL, but no program position at all is taken on the lower limits of, say, unilateral PHL that justify candidacy for amplification.

### Protocol Challenge, Improvement and Adaptation

A good protocol is not written in stone. It does not change on a dime, but it can and should be amenable to evolution and also to adaptation. Evolution is inevitable because the evidence base changes over time, experience is accumulated, new insights dawn and lessons are learned from quality improvement efforts, new technologies are introduced, and so on. However, even if none of these things happened, the protocol should be able to change. When developing a program protocol it is impossible to think of everything. Protocol developers themselves will make errors and omissions (preferably minor) that will come to light in clinical practice, espe-

cially in large EHDI programs. Who better to be able to initiate change than the very people who are obliged to use the protocol? Audiologists should be actively encouraged to challenge the protocol and to suggest deficiencies, things that don't seem to work and areas that have been overlooked. Such input should be treated seriously and respectfully. Options include new evidence reviews, surveys of all program audiologists or at the very least a cogent and constructive response to the challenge or suggestion. It should not be ignored or dismissed.

It is easy to see how the process just described fits with sound principles of knowledge translation and also relate to several of Cabana and colleagues' (1999) barriers discussed earlier. Engagement of the recipients of a protocol in a genuine dialogue is likely to encourage a sense of participation and ownership, as well as reducing any sense of loss of control and abrogation of professional freedom of practice. Of course, it is to be hoped that the "user community" of any protocol would have had an opportunity to contribute in a genuine and well-structured way to the original development of any protocol. There is nothing quite like inclusion as prophylaxis for alienation.

An important distinction must be made between protocol evolution and adaptation. Evolution occurs over time and adaptation (at least, as the term is used here) occurs over space or practice contexts. In order for audiologists to be able to adhere to a given guideline or protocol, they must be in a position to do so. Such a position implies a measure of control over the practice environment, procedures, scheduling and so on, or at least endorsement of protocol imperatives by controlling individuals, agencies or institutions. An obvious potential issue is the additional resources that may be necessary in order to adhere to a protocol that is determined externally to a given practice context. The drive for quality that may be appropriate at the program-wide level may far exceed that which is the norm for any individual practice context. Even if that is not the case, resources of test time, test environment, instrumentation and infrastructure such as data systems, administrative and clerical support, etc., may be in short supply. This raises the issue of asking someone to do something without the support needed to do it. While human inventiveness is a powerful force if driven by conviction, there are limits to what can be accomplished. What must be recognized is that a protocol has resource consequences of many kinds and this should be planned for and accommodated programmatically. Furthermore, there may be a trade-

off between the resource or change required in a given practice context and the extent of protocol adherence that is practicable. The bottom line is that half a loaf is better than nothing, and this reality requires a very clear view of the onion that is protocol adherence. Protocol elements can be grouped in layers of importance and the question is: under a given set of real practice constraints, as distinct from artificial constraints designed to avoid practice change or to inflate budgets, which protocol elements may be given up and which must be retained?

### Clinical Decision and Protocol Support

It is not enough simply to make a protocol available. Effort is required to encourage its use, to explain it, to enhance it and to generally make it as accessible and relevant as possible to daily practice. There are many ways to do this. The IHP supports audiologists who provide diagnostic and amplification services in several ways. In diagnostics, for example, this includes making available expert guidance on questions or issues of protocol content and interpretation, consultative review of individual test results or entire cases, advice on interpretation of individual test results, reporting and follow-up test strategy. Protocol revisions or addenda are issued as and when new information comes to light from the scientific or clinical literature, from program policy decisions or as a result of accumulating program data and audiologist experience.

### The Bottom, Bottom Line

Good protocols are absolutely and unquestionably essential for all major components of a high-quality EHDI program. Without them, high effectiveness, equity and efficiency of services are virtually impossible to achieve, sustain or improve. Good protocols are difficult and costly to formulate, disseminate, maintain, support and verify. The true health professional will welcome a good protocol as a repository of current evidence and clinical insight, as well as a bulwark against significant clinical errors. Good protocols take full account of the uniqueness of the individual child and family, as well as of underlying commonalities of logic and biophysics. They should not overly constrain professional freedoms of practice but should be a judicious blend of discretion and requirement. They should facilitate the direction of individual clinical expertise, judgment and effort into practice areas of limited evidence or of such complexity or subtlety that they cannot be addressed by simple

---



rules, however well-founded. Most importantly, a good protocol should not only state exactly what should be done, but why, when and exactly how it should be done. Audiologists involved in EHDI services delivery should be advocates for nothing less.

## References

- Agency for Healthcare Research and Quality (AHRQ). Retrieved 10 May 2011 at: <http://www.ahrq.gov>.
- AGREE Collaboration. Retrieved 10 May 2011 at: <http://www.agreecollaboration.org>.
- Andermann, A., Blancquaert, I., Beauchamp, S., and Dery, V. 2008. Revisiting Wilson and Jungner in the genomic age: A review of screening criteria over the past 40 years. *WHO Bulletin* 86(4): 241–320. Retrieved 10 May 2011 from: <http://www.who.int/bulletin/volumes/86/4/07-050112/en/index.html>.
- Armstrong, R., Waters, E., Roberts, H., Oliver, S., and Popay, J. 2006. The role and theoretical evolution of knowledge translation and exchange in public health. *Journal of Public Health (Oxf)* 28(4):384–9.
- Cabana, M.D., Rand, C.S., Powe, N.R., Wu, A.W., Wilson, M.H., Abboud, P.A., and Rubin, H.R. 1999. Why don't physicians follow clinical practice guidelines? A framework for improvement. *Journal of the American Medical Association* 282(15):1458–1465.
- Center for Evidence-Based Medicine (CEBM). Retrieved 10 May 2011 at: <http://www.cebm.net>.
- Cochrane Collaboration. Retrieved 10 May 2011 at: <http://www.cochrane.org>.
- Donabedian, A. 1990. *The seven pillars of quality*. *Archives of Pathology and Laboratory Medicine* 114(11):1115–1118.
- Francke, A.L., Smit, M.C., de Veer, A.J., and Mistiaen, P. 2008. Factors influencing the implementation of clinical guidelines for health care professionals: A systematic meta-review. *BMC Medical Informatics and Decision Making* 8:38.
- Gravel, J. 2005. Evidence-based practice in pediatric audiology. In R. Seewald and J. Bamford (eds.), *A sound foundation through early amplification 2004: Proceedings of the third international conference* (pp. 17–26). Stäfa, Switzerland: Phonak AG.
- Grosse, S.D., and Ross, D.S. 2006. Cost savings from universal newborn hearing screening. *Pediatrics* 118(2): 844–845.
- Hyde, M. 2005. Evidence-based practice, ethics and EHDI program quality. In R. Seewald and J. Bamford (eds.), *A sound foundation through early amplification 2004: Proceedings of the third international conference* (pp. 281–302). Stäfa, Switzerland: Phonak AG.
- Hyde, M. 2011. Principles and methods of population hearing screening in EHDI. In R. Seewald and A.M. Tharpe (eds.), *Comprehensive handbook of pediatric audiology* (pp. 283–338). San Diego: Plural Publishing.
- Joint Committee on Infant Hearing (JCIH). 2007. Year 2007 position statement. Principles and guidelines for early hearing detection and intervention programs. *Pediatrics* 120:898–921.
- Lipsey, M.W., and Wilson, D.B. (2001). Practical meta-analysis. *Applied social research methods series* (Vol. 49). Thousand Oaks, CA: SAGE Publications.
- Straus, S.E., Tetroe, J., and Graham, I. 2009. Defining knowledge translation. *Canadian Medical Association Journal* 181(3-4): 165–168.
- Ward, V., House, A., and Hamer, S. 2009. Developing a framework for transferring knowledge into action: A thematic analysis of the literature. *Journal of Health Services and Research Policy* 14(3):156–164.
- Wilson, J. M. G., and Jungner, G. 1968. Principles and practice of screening for disease. Geneva, World Health Organization. Public Health Paper # 34. Retrieved 10 May 2011 from: [http://whqlibdoc.who.int/php/WHO\\_PHP\\_34.pdf](http://whqlibdoc.who.int/php/WHO_PHP_34.pdf)

