THE MAIN ARGUMENT SECTION

Cochlear Implants – the Sentence Test Fallacy

CONTENTS

Copyright © Nicholas Dollar 2005

Professor J T Rubinstein, MD, PhD
Dept. of Otolaryngology, Physiology
& Biophysics, Biomedical Engineering,
The University of Iowa
200 Hawkins Drive
Iowa City
IA 52242
USA

10th January 2003

Dear Professor Rubinstein,

Cochlear Implants – the Sentence Test Fallacy

In October you kindly sent a copy of your paper "Residual Speech Recognition and Cochlear Implant Performance: …". I have read it with fascination and, for reasons that shall become clear if you have the patience to read this letter, I cannot resist raising a number of issues with you in the hope that you shall be prepared to respond.

Putting it in nutshell, I invite you to consider whether using sentence test scores in your formulae misleads rather than clarifies.

I am a candidate for implantation and a lawyer, not an audiologist, so I would not expect you to take what I say seriously without my explaining myself, and I am not familiar with the shorthand of your profession; I hope this passes as an excuse for my verbosity.

I start by explaining my background, and the genesis of my fundamental concept that a sentence test is a wholly unreliable measure of a patient's hearing. It is a concept continually validated by my own experience, but I have found some support for it from audiologists and some statistical corroboration of the test's inaccuracy. I then suggest a number of ways in which an implant might, in theoretical terms, affect hearing and how it might be measured. I point out that the sentence test seems to have acquired a dominance that I suggest is unjustified, being a yardstick used almost universally, so far as I am aware, when candidacy for an implant is to be determined. I wonder whether its dominance is such that it has driven you to act like some pre-Copernican astronomer, devising mathematical formulae to account for the movements of the stars on the surface of a sphere that he failed to recognise, or was compelled to deny, was fictitious. I then attempt to explain how the formulae in your paper might take us further away from understanding the processes they seek to elucidate.

I am quite prepared to make an utter fool of myself, but, having struggled in vain for about a year to persuade my Programme to provide a coherent response to my queries, any reply from you would help me to make sense of the system. Many years ago I concluded that audiologists were an earnest and resourceful group of people valiantly exploring virgin territory armed with an empirical compass but no map (they were unable to predict which of a choice of aids would help me most, and left me to choose whichever I was most "comfortable" with). I had expected the advances in knowledge and the development of implant technology to change that, but I am yet to be convinced.

I am 53 years old and experienced what might have been a congenital, but certainly was a progressive, hearing loss so that my PTA loss was about 55dB by the age of six (from which age I have been a hearing aid user) and passed 90dB by the time I was 25. It continued to increase at a constant rate before the high frequency loss fell off the chart but, assuming the same rate, it is now about 138dB. This probably gives me a net loss for my aided hearing of about 88dB, although I have no apparent high frequency thresholds. I have been turned down for implantation because my score in the "open set speech discrimination test" (i.e. sound-only sentence comprehension test) was 74%, which greatly exceeds the 30% maximum score required here for eligibility. I was told that the average score for post-implant patients is about 60% and that there was therefore a real risk that I might lose performance.

That reasoning seemed manifestly absurd, because I knew that I had *heard* perhaps only one quarter of the sound elements in the 74% correct responses; the rest I had filled in using intelligent guesswork or what I now call supplementary sound interpretation. Enclosure 1 is a copy of my letter of 20th Jan '02 to my Programme in which I explain my view of how *hearing* is enhanced by supplementary sound interpretation, which is a cerebral, not a physical, activity, and that a sentence test is measuring the level of comprehension produced by the combination. I would add two points about the letter.

The distinction I draw between "skeleton" and "reflective" interpretation can best be illustrated by a homely example. My wife used to find irritating my frequently responding to a question by saying "Pardon", and my then answering it before it was repeated. Quite understandably, she saw this as a verbal "tic", but it was not. In normal conversation, one responds immediately, and my immediate response was that I had not heard the question. However I would re-play what I had heard in my mind and often I could solve the puzzle before it was repeated. What I had *heard* was not enough to understand the question using skeleton interpretation, but was sufficient to act as the basis for a process of interpretation akin to a computer OCR program incorporating context and probabilities. The extreme form of this process occurred when my wife unsuccessfully repeated the question twice, lost patience and said, "Forget it!", and I would exclaim after several minutes' further conversation on another topic, "Oh, I know what you were asking.". The process of supplementary sound interpretation had finally yielded a result.

The second point relates to my graph in Table "C". You will note that, rather than follow my instincts, I cautiously assumed supplementary sound interpretation ("SSI") could account for no more than 66.6.% of the score in a sentence test. I have since found some statistics that seem to indicate my instincts were correct. If one assumes that a monosyllabic word test accurately measures the acoustic information a candidate receives from speech (i.e. *hearing*), then the figures in Enclosure 2 suggest that SSI can account for up to 74% of a sentence test score. You will note that Professor Dorman states that the patients who scored 90-100% in a sentence test scored an average of 58% in a word test. He mentions the best performance in a word test was 90%, 32% in excess of the average, so, if one assumes the results were evenly spread, then one can conclude the worst performance was 32% less than the average, namely 26%. Naturally, without having all the figures, any conclusion is partly speculative, but, on the face of it, the true hearing of the group who scored 90-100% in the sentence test ranged from 26% to 90%. The contribution of SSI might have ranged from 74% [100% in sentence test minus 26% pure hearing] to nil [90% in sentence test minus 90% pure hearing].

If it is correct that the word test accurately measures *hearing,* then it must follow that the direct impact of an implant is on hearing as measured by the word test. There are several possible scenarios. Because the front line of the pre-operation hearing receptors are replaced by the implant, it might be that the outcome is constant regardless of pre-operation hearing, or produces a result over a range reflecting only the ability of the technology to be consistent. I assume the technology has been tested to assure a consistency within narrow parameters, so inconsistency of technology cannot explain the wide range apparent from Professor Dorman's figures. So, if one looks for an element of positive or negative accrual, relating pre to post-operative hearing, one has to look elsewhere. Could the range be accounted for by inconsistencies in the second line of hearing receptors, namely the spiral ganglion neurons? Even without considering the issue of neuronal deterioration, the implant, as I understand it, is capable of stimulating only a fraction of auditory nerves that would be available for use to a hearing person. One could speculate that, in a hearing person, a proportion of the auditory nerves do not function and are otiose because of the overwhelming preponderance of functioning nerves. A miniscule alteration in the positioning of the implant might have a disproportionate effect on the number of such non-functioning nerves relied on. If that were to account for the wide range of outcomes, however, the outcome for any given patient would be inherently unpredictable, or certainly so with the tools at present available.

So if there is predictability of outcome, it seems reasonable to attribute it to the quality of the second rank of hearing receptors on the basis that, given wholly undamaged second rank receptors, the optimum outcome could be achieved, representing the full potential of the implant. Naturally, one cannot know the optimum outcome because no "ceiling" has yet been reached and no patients with perfect hearing have been implanted, so the optimum remains a notional and unknown figure of x% score in a word test. Of course it may be that the optimum is beyond the range of the word test; it might, for example, be a score of 200% in a stiffer test with the same calibration. The degree to which the actual outcome falls short could be a reflection of the deterioration of the second rank, but, since that deterioration cannot be measured directly, one must look for outward signs of it.

Perhaps the most temptingly simple correlation to seek would be that between pre-operation *hearing* and the capacity of the second rank receptors. You mention in your paper the unsuccessful attempts to relate pre-operation pure-tone audiograms to post operative performance, but I have found no reference in your paper or elsewhere to the use of word tests as an indicator of the health of second rank receptors. It was a reference to Professor Gantz' paper of fourteen years ago[1] describing his team's use of such tests as part of a dual qualification for eligibility that prompted me to contact your institution, so I am puzzled that you have not made more use of them. In stead, you choose to rely on sentence tests as a guide to pre-operative capacity.

While I can understand why sentence tests were originally adopted as the key measure of pre and post-operative performance, I am mystified by their continuing dominance in the cochlear implant field, and in particular by their almost universal use as a determinant of eligibility. Enclosure 3 is a copy of an email I sent to Professor Dorman in which I set out my views more fully and explain how I fear some clinicians might have lost sight of their proper clinical aims because they are unwittingly using as if it were a measure of *hearing* a test that primarily measures a candidate's adaptation to hearing loss. Even if one accepts the questionable premise that some system of rationing is required because of the relatively high cost of implantation, using sentence test scores as a qualification makes invidious choices even more so. Under the present fiction that a sentence test score accurately measures hearing, one can see why a candidate whose score is anticipated to rise from 30% to 60% would be preferred to one whose score is anticipated to rise from 40% to 60%; value for money is combined with helping the candidate more severely affected by hearing loss. If the fiction were discarded, however, the implicit value judgements would need to be re-assessed. Which of two candidates with *hearing* of 14% (as measured by a word test) should be preferred, the one whose performance in a sentence test is anticipated to rise from 14% to 60%, or the one whose score might rise from 42% to 100% (or 180% in a stiffer test with the same calibration)? The second candidate clearly offers better value for money but with a cut-off point of 30%, as in the UK, or the FDA's 40% in the US, he is not even considered at present.

Given this background, you will understand that one of my first thoughts on reading your paper was that medical politics might have influenced your attempt to relate two quite different measurements of hearing; that you felt obliged to adopt sentence test scores as a pre-operative starting point because you knew they were used to determine eligibility but that you were trying to move your colleagues away from over-dependence on them by expressing the success of implants in terms of word test scores. Whether or not you are a closet Copernican, I hope you will bear with me if the points I make below strike you as being too obvious.

Professor Dorman's figures illustrate that word and sentence test scores are measuring different values – as different as weight and volume. Even if one reaches no conclusion about

[11] GANTZ, B.J. (1989). Issues of candidate selection for a cochlear implant. *Otolaryngol. Clin. North Am.* 22(1), 239-247.

which more truly gauges *hearing*, assessing the impact of an implant by relating the pre-operation figure for one to the post-operation figure for another must, in principle, bear risks of misleadingly suggesting a connection that is apparent rather than real.

Your formulae also use duration of deafness, measured in years since the candidate abandoned the use of the telephone. Might I mention how subjective this measurement is and that it must, to some extent, reflect the candidate's lifestyle, dependence on the phone, access to special aids and willingness to persevere? Overwhelmingly however, it must reflect the degree of SSI displayed by the candidate. Enclosure 4 is a transcript of an amusing phone call I received and it shows the extent to which SSI plays a role. By the time I had established that it was an unsolicited business call at my home on a Saturday, the caller being a double-glazing salesperson was probably at the top of a quite short list of possibilities; I do not know what the equivalent might be in your part of the States, but here double-glazing firms are the most annoyingly persistent of tele-sales adherents. My point is that candidates with high levels of SSI are likely to have a shorter "duration of deafness" by your measure and that, only in the case of candidates who suffered a rapid major hearing loss and immediately stopped using the phone, would "duration" reflect time elapsed since *hearing* had deteriorated below the point at which it alone could sustain use of the phone.

[NOTE FOR READERS OF THIS SITE: "CID" is a sentence test, and "CNC" is a word test. See Prof Rubinstein's paper for fuller definition]

A further general point to be made is about the sample of patients whose data was analysed. Because of the nature of the selection procedure, the patients are not a representative cross section of those with very limited hearing. Since the FDA prohibits implants for those scoring more than 40% in a sentence test, your sample will have a bias towards those with lower levels of SSI, the bias being greater the higher the CID. I am not sure whether the dual criteria of word and sentence test scores used at your institution are mutually exclusive; I was slightly puzzled to be told by Professor Gantz in his e-mail of 28th September that you were implanting patients with words test scores of about 20% - I had assumed you would treat candidates with a score in such a test of up to 40% provided they also fell within the sentence test limit i.e. you would treat patients with CNC of 40% and nil SSI. If you do use 20% CNC as a separate cut-off, then the bias towards those with lower SSI would be reduced but not eliminated. Since SSI can contribute up to 74% towards CID scores, there will still be candidates who pass the CNC hurdle but fall at the CID (e.g. 20% CNC but 77% CID).

(It might of course have been that you implant patients with CNC scores up to 20% regardless of their CID score, but not only would this breach the FDA embargo, I note that your formula 2 has not had to cater for candidates with nil "duration". Someone with hearing as in the example at the end of the last paragraph would have hearing similar to my own, and would no doubt continue to use the phone as I do.)

Your formula 1 provides that, for any patient, his post-operative word test score shall be 53.9, increased by 13% of his pre-operative sentence test score, and then reduced by between 42% and 119% of the duration of his deafness in years. The higher his pre-operative sentence test score, the lower shall be this reduction.

Because the CID score can be such a misleading indication of true hearing, it must be the case that your formulae only work to the extent they do because you have arrived at a working average between the two extremes of SSI displayed in your sample. That "averaging" can be seen at work if one compares the results calculated using the third and fourth lines of formula 2 with those arrived at using formula 1. In both cases, formula 1 overestimates in the lower ranges and underestimates in the higher ranges of CID. (This does not apply to the first two lines of formula 2) Perhaps the sliding scale, or a significant proportion of it, applied to the reduction for duration in formula 1 is only required to compensate for the increasing bias towards lower SSI the higher the CID score in your sample. In other words it is, as in my view common sense would suggest, residual *hearing* as measured by CNC (combined with duration) that is the determinant, not residual CID.

I am not suggesting that duration of deafness does not result in lower outcomes, but I do wonder whether using pre-operative CID as if it were a determinant is misrepresenting the damaging effect of duration. It seems to me quite possible that it is not long-term low residual hearing *per se* that leads to lower outcomes, but long-term lack of use of such residual hearing.

It is not even as if you are using CID to assess duration; you time that not from the patient's hearing dropping to a particular level of CID, but from his subjective decision to abandon using the phone Even if this were to happen at a relatively uniform level of CID, the level of CNC at the point the decision is made must vary enormously. It must, surely, be the level of CNC that reflects the health of the spiral ganglion. It might be that someone with higher SSI is more vigorously exercising the remaining healthy ganglion because their heightened speech reception encourages more frequent reliance on hearing. This might in turn delay the onset of deterioration that might only set in when residual hearing is no longer exercised. Perhaps it is the presence of an actively used part of the ganglion, rather than the mere size of this part, that is significant. This is all speculation, but to suggest that the outcome of an implant measured in terms of CNC can directly reflect a purely cerebral activity like SSI seems to me to almost metaphysical in conception. It seems almost as far-fetched as attempting to include lip-reading skills in your formulae.

So, why make these compromises and intricate mathematical adjustments to bridge two quite different measurements – to predict the resulting weight of an object subjected to a process of incrementation when all you know is its original volume? You say in your paper that you have chosen to do so to avoid "floor" and "ceiling" effects. Please correct me if I am wrong, but I take it that this refers to the phenomenon that a significant proportion of your present patients start with CNC scores at or too close to zero to be satisfactorily distinguished, and that a similarly sizable proportion emerge from implantation with CID scores bunched close to the 100% figure. Unless I have misunderstood this, would the better solution not be to develop more sophisticated methods of measurement, rather than continue to relate two different values in a way that continues the myth that sentence tests accurately measure hearing?

I can hardly disguise the strength of my feelings about the sentence test fallacy, which have grown over the months of my correspondence with my programme. Their obtuse and evasive replies have left me wondering with bewilderment whether anyone in the cochlear implant field has considered exactly what they are observing when they see a difference between word and sentence test scores. Beyond vague references to the use of "sentence context" and "top-down processing", nobody has referred me to a theory that accounts for the difference. In particular nothing has been said to me, nor have I found anything, that contradicts my fundamental concept that sentence tests are a wholly unreliable measure of hearing, and that SSI is a cerebral, not a physical, activity.

Please accept my assurance that no offence is intended despite what must seem to be my provocative and uninformed effrontery. I am eager for enlightenment, so I very much hope you shall feel able to respond.

Yours sincerely,

N M W Dollar

Letter to my Programme Co-Ordinator (Consultant Audiological Scientist)

Xxxxxx xxxxxxxx
xxxxxxx
xxxxxx
xxxxxx
xxxxxx
xxxxxx

20th January 2002

Dear Mr xxxxx,

You will be aware that I wrote to Mr xxxx on the 16th December and that he kindly dictated a reply on the 19th indicating that you would be writing dealing with some of the questions I raised. I am very aware that Christmas and the New Year have intervened, but I would be grateful to hear from you soon. Meanwhile what I set out below may help you understand more fully the nature of my concerns.

Just as lip-reading skills contribute to communication, and can be regarded as a constant factor before and after an implant, so too, I have suggested, do sound interpretational skills. The way I approach it, there are three kinds of interpretation involved that could be called “linguistic”, “skeleton” and “reflective”.

“Linguistic” interpretation is almost synonymous with linguistic ability; it detects the pattern of sounds that form words in a particular language. To a hearing person embarking on learning a foreign language, fluent speech in that language would initially seem like an uninterrupted flow of unintelligible gibberish. Everything is heard but, until some basic knowledge of vocabulary and grammar is acquired, even the gaps between sentences cannot be detected, let alone those between words. As the learner becomes increasingly familiar with the language, it is easier to recognise patterns and understand what is said. This is of course characteristic of any language including BSL but we are only concerned with sound-based language. The greater one’s knowledge of the language, the more patterns are instantly recognisable as a meaningful interpretation of the sound produced by the speaker.

“Skeleton” interpretation involves recognising a shorthand version of the patterns. Given a thorough familiarity with the language, even a hearing person would have little difficulty recognising certain distinctive patterns in circumstances where some sound elements are missing, but where there are contextual clues available. If in Court defence Counsel stood up and said “***ah***-ah-ah-***ah***-ah”, everyone would be mystified. An audience watching a magician flourish his magic wand over his inverted top hat, saying “***ah***-ah-ah-***ah***-ah”, would almost certainly know instantly that “Abracadabra” was the word. A deaf person, assuming he hears some of the speech frequencies, uses this second kind of interpretation all the time. It is a matter of instantly recognising the skeleton pattern as a substitute for the fleshed-out pattern a hearing person perceives.

“Reflective” interpretation arises where there is no instant recognition of the patterns and it is necessary to reflect on what has been heard, or have it repeated, before an intelligent guess can be made. This interpretation naturally uses contextual clues but, in contrast to "skeleton" interpretation, the clues need to be teased out because the skeletons are too ambiguous or perhaps have limbs missing, because too many sound elements are missing.

As a concrete example of these forms of interpretation at work, one of the sentences in the tests I took was along the lines, "Children suffer illness more frequently when they first start school". As I recall it, recognition of the sentence was in this order – "school-children-illness-more". You will note that each word leads on to, and then is reinforced by, the next. With this information I guessed the thought being expressed which is particularly commonplace for me having had five children and I was then able to fill in the remaining words to make sense of the sentence. (It would surprise but not astonish me if you were now gently to point out that this was a sentence I got wrong! I am constantly flying by the seat of my pants and a misinterpretation of a sentence like this would normally become evident from the flow of conversation.)

If an implant were capable of producing perfect hearing, then the supplemental forms of interpretation -"Skeleton" and "Reflective"- would be otiose after the operation, as would lip-reading skills. The patient may however have to learn or re-learn linguistic interpretation if his previous experience of speech was beyond his memory or had been non-existent or very limited. In reality, it seems that an implant is certainly not capable of producing perfect hearing, so both the forms of supplemental sound interpretation would continue to play a role. Depending on how effective the implant is in supplying previously missing sound, interpretation may still be fully stretched but more fruitful, or might at times be unnecessary. The skeletons would have at least some flesh and probably no missing limbs.

Interpretational skills, as with any other, need to be acquired through learning and practice, so one would not expect them to be present in equal measure for all deaf people. Someone who is highly articulate has both breadth and depth of linguistic knowledge and a sure touch that enables him to perform gymnastics with the language showing an agility not given to others. Such a person, who almost by definition would have a high level of linguistic interpretation, would be at a distinct advantage when it comes to Skeleton and Reflective interpretation. Conversely, someone with poor linguistic interpretation would have difficulty instantly recognising as meaningful any but the most common words even with full hearing, let alone when they depended on Skeleton or Reflective interpretation. So, Linguistic interpretation is one variable.

Similarly, one would expect experience to improve the supplemental forms of interpretation. An adult with many years of full hearing, who quite rapidly descends to profound deafness, would initially display almost no interpretational skill and may find the process of developing it harder and more prolonged than a similarly affected child would. Whatever the background, it is reasonable to assume that some people would cultivate the skills more diligently or successfully than others.

How large is the contribution of the supplementary forms of interpretation? In everyday life I frequently create situations where it is almost unnecessary for me to hear any part of what is said to me because I have said something the response to which I can confidently predict. The merest hint of the expected sounds confirms my prediction. Here, interpretation probably supplies 95% of understanding. In the artificial circumstances of the sentence recognition test, I had no opportunity to define parameters but, even there, interpretation played a formidable role. Taking the example of the sentence mentioned earlier, about schoolchildren's illnesses, I probably heard less than one quarter of all the sound elements involved. This may seem startling, and for me to make any kind of accurate assessment is immensely difficult, but I draw some reassurance from my calculating in theory that the PTA loss for my hearing, using aids, falls within the category of "severe" loss, i.e. 71-90 dbs.[2]. You took an audiogram of my aided hearing but did not mention the result; it would be interesting to know. If my aided hearing is at this level, perhaps the figure of one quarter is a significant over-estimate.

Even if one took a more cautious figure, and assumed that, in the test, sound contributed one third rather than under one quarter to my performance, the graph in table "C" shows clearly where my concern lies. The vertical axis represents the score in the test. The horizontal axis is the candidate's hearing – 100% being an ability to hear, with an aid if used, all sound elements produced by normal speech. [I am ignoring my view that the conditions of the test are so favourable that it is almost as if an additional hearing aid had been supplied, and am treating a score of 100% in the test as a measure of 100% hearing]

I have assumed a direct correlation between hearing and performance, hence the straight line rising diagonally at 45° across the graph from 0% to 100%, although there must be various doubts about this. For someone with perfect hearing, the first 5% loss may be relatively more significant to understanding than the next 5%, so it might be that there should be a steeper incline as one approaches 100%. Alternatively, for someone with no hearing, acquiring some hearing would represent an infinite improvement, so perhaps the incline should be steeper close to zero. A further alternative is that only above a certain critical point does hearing enable any usable recognition of speech frequencies, so the incline should be shallow at first and then rise more steeply. I have no idea which of these approaches, if any, is correct but I doubt whether any of them would invalidate the point I am seeking to make.

Above the diagonal "hearing" line I have added a delineation of the combined contribution made by Skeleton and Reflective interpretation at its maximum. For the purposes of demonstrating my point, I have assumed that the conservative estimate of a two-thirds contribution to understanding in my case is the maximum achievable, although I have no doubt there are many people with a greater capacity than mine. I have shown it as a straight line, although similar doubts could be expressed about this as have been mentioned about the "hearing" line.

Several points become obvious from the graph: -

1. A score of 75% could represent the result from a candidate whose hearing is 25%, as it is for me, or 75% from one who had developed no interpretational skills. (See box "A-A-A-A")
2. The benchmark score of 30% could apply to candidates with hearing of 30% but no interpretational skill, in other words candidates with more hearing than mine. (See box "B-B-B-B")
3. The 60% average score of post-implant patients could represent hearing of 60% for a patient with no interpretational skills. (See box "C-C-C-C") If it is in any way a valid prediction of outcome, and an implant were to provide me with hearing of 60%, I would easily score 100% in the test. I would need only to improve my hearing to 34% to achieve a perfect score. (See "D-D")

[2] Current PTA loss approximately 136 dbs (see table "B" with my letter to Mr xxxx of 3rd Oct 01). Deducting probable amplification provided by aids of 50 dbs leaves PTA loss of aided hearing at approximately 86 dbs.

This is all grossly over-simplified, but I hope you see why I have serious doubts about the prominence awarded to the sentence recognition test.

The key questions for someone in my position are, firstly, whether an implant is likely to introduce useful sound in frequencies now unavailable to him and, secondly, whether that gain would be at an unacceptable cost in terms of loss of quality of sound in the frequency ranges he does now hear. Contrary to any impression I might have given, I am not determined to have an implant come what may, what I had expected was an assessment process that provided at least some of the information needed to answer the key questions.

In stead, I feel I have been an unwitting participant in a hopping race. If there were ten one-legged patients in the waiting room of an orthopaedic hospital, and only four prosthetic legs, there would be a certain crude logic in the staff organising a race through the grounds. The first six across the line could be awarded certificates of merit – and sent home. The staff could then concentrate on helping the remaining four, most disadvantaged, patients although the staff may feel more comfortable wearing "Race Marshal" badges rather than those of medical rank. Certainly the patients sent home would feel less frustrated if they had been told at the outset that whether they would receive help would depend on an arbitrary test of "need" rather than potential benefit. It would be an insult to tell them it was all for the best because most of them covered the course faster than the average for patients fitted with an artificial leg. (The last man in the race did not leave the start – he has lost all sense of balance and cannot stand even with two legs)

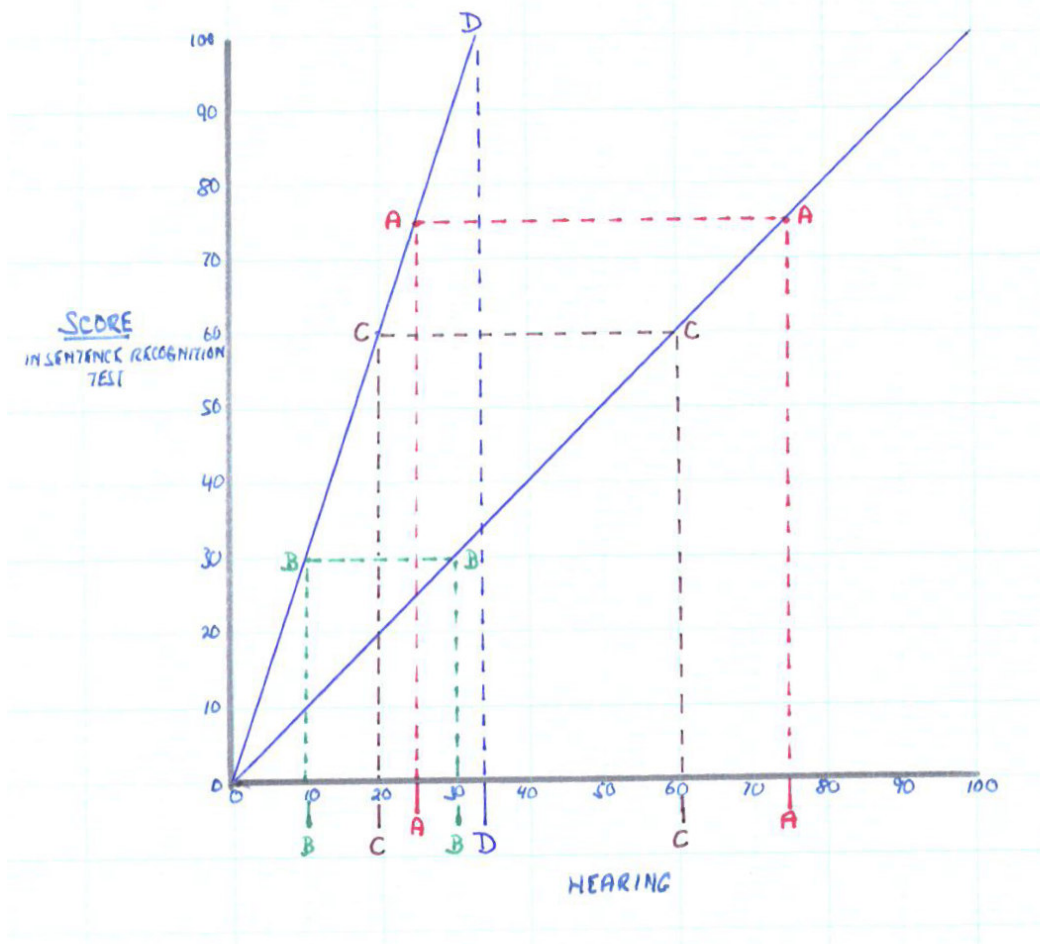
This is of course an unfair analogy. It might be closer if the only way of treating the patients was to remove the good leg and fit two artificial ones. The point remains, however, that selection is arbitrary rather than based on a careful assessment of potential benefit.

I enclose a copy of a letter I am today sending to Mr xxxx, and have sent him a copy of this letter. This may well be quite unnecessary, but I hope it might save some time and avoid copies having to be passed around the hospital. I look forward to hearing from you as soon as possible.

Yours sincerely,

N M W Dollar

TABLE 'C'



D
A
A
C
C
B
B
SCORE
IN SENTENCE RECOGNITION TEST
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
B
C
A
B
D
C
A
HEARING

From **Cochlear Implants: A Handbook** © 1998 Bonnie P Tucker by permission of McFarland & Company, Inc, Box 611, Jefferson NC 28640 www.mcfarlandpub.com

Chapter 1 : "An Overview of Cochlear Implants" by MICHAEL F. DORMAN, Ph.D.

Michael Dorman received a Ph.D in Developmental Psychology from the University of Connecticut. He is currently a professor in the Department of Speech and Hearing Science at Arizona State University and an adjunct professor in the Department of Surgery, Division of Head and Neck Surgery/Otolaryngology at the University of Utah Health Sciences Center. He directs research programs for individuals with cochlear implants at both Arizona State University and the University of Utah Health Sciences Center.

Pages 5 up to and including most of page 17 omitted.

Page 17 - **The Current State of the Art**

How well do patients understand speech with the current generation of cochlear implants? Scores for words in sentences for patients who use the eight-channel Clarion CIS processor are shown in Figure 8. The most common score is 90-100 percent correct. Similar scores are found for patients who use Cochlear Corporation's Nucleus 22-Spectra processor, which is the most widely used processor in the world (with over 15,000 units worldwide). However, for any device used, some patients, perhaps 10 percent, score in the 0-10 percent correct range. As we noted in the section on anatomy and physiology, inevitably some patients will have very few fibers to stimulate and, as a consequence, will not be able to code the multiple frequencies in the speech signal.

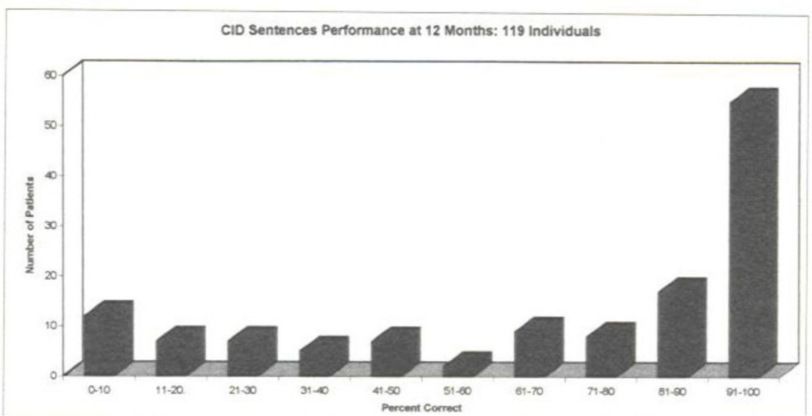


Figure 8. Identification of words in sentences by 119 adults fit with a cochlear implant. Slightly more than half of the individuals achieved scores between 80 and 100 percent correct.

Even the best-performing patients do not get all the information in the speech signal. Patients who achieve scores in the sentence test of between 90-100 percent correct average only 58 percent correct when tested with individual words without sentence context (the best-performing patients achieve scores between 70 and 90 percent correct). Thus, it appears that the best-performing patients recognize many, perhaps most, individual sounds and rely on sentence context to fill in the identity of the sounds they cannot resolve.

Email to Prof Dorman

Dear Professor Dorman,

Many thanks for your reply which, again, is refreshingly sane and helpful. I am most grateful for your offer to enquire about tests for candidates and look forward to hearing further.

From your travel plans, it looks as though you might be attending the "7th International" in Manchester. (*Rest of paragraph omitted*)

While you are over here, if you have any opportunity to visit Devon I would be delighted to take you, and your partner if you are accompanied, for a meal. Do please let me know. If you haven't seen it before, you will find the West Country is one of the most delightful parts of the UK.

Meanwhile, I am afraid I cannot resist the temptation to leave you with some food for thought. Since you clearly take the same view as I do over the sentence test, that it measures ability to cope rather than hearing, I can quite understand your also taking the view that the Establishment is *intentionally* rejecting those better able to cope. If I am to buck the system, it follows that I must either find someone prepared to ignore the "rules", or bamboozle the system into thinking I fall within them. On one level this must be a correct description of the practical reality both in the UK and the US. I am not yet sure, however, that such a pragmatic and rather cynical analysis embraces the whole story.

There must frequently be tension between the imperatives of the funding and the clinical arms of any health system. The clinical arm would aim at a complete "cure" of a physical disability, so a clinical judgement on the merits of any given treatment would (or certainly should) be based on whether, on balance, the treatment is likely to provide the cure, or a sufficiently significant step towards such cure, with reasonable certainty and acceptable risk. The funding arm would be concerned with looking at the situation from the other end; rather than aiming upwards towards a cure it would look downwards at how low the level of function could be while remaining "acceptable" in the sense of not justifying expenditure. It is a cliché that what is "acceptable" is notoriously flexible and almost infinitely debateable, and, at the end of the day, the way any given health system makes such decisions depends on a complex mixture of factors including how far it is pushed by the clinical arm and how responsive it is to "public expectations". Public expectations can be heavily influenced by the mere knowledge that the clinical arm has available what is, or approximates to, a cure.

So, in essence, the clinical arm has an absolute and objective standard, while the funding arm has a subjective one. As a matter of practical medical politics, naturally the bodies responsible for funding try to draw in members from the clinical arm and reach an accommodation with them, but if the clinical arm lost sight of its different perspective, the result would be stagnation. The benefits made available by advances in medical science would only see the light of day in so far as they provided a more economical or efficient way of assisting those falling below the funding arm's view of an "acceptable" level of function. That might take *them* closer towards a cure than before, but would fail to benefit those for whom a "cure" is now available but who had previously been seen as having an "acceptable" level of function partly because it was not known a "cure" was available.

I fear that the sentence test fallacy has so deeply infected the Cochlear Implant establishment that they have lost sight of their proper clinical aims, and have played into the hands of the funding arm by willingly adopting as their key measure one that, rather than focussing on the physical disability, concedes any argument there might be over what is an "acceptable" level of function. It may well be that the origins of the test were entirely innocent and well intentioned. In the early mid-years of implantation, when the technology had passed the stage of being so crude that the sacrifice of *any* useable speech hearing capacity could not be justified, the hearing figures must have been small, so it did not matter much if pre and post-implant hearing was not measured with atomic precision. The sentence test was convenient and provided a simple way of gauging the effectiveness of implantation. Now the figures cannot be described as small, so any mismeasurement of *hearing* implicit in the test has grown proportionally to such an extent that the test is almost meaningless as a measure of hearing. Yet it still seems to dominate clinical thinking, which in principle should be focussed on the physical disability for which a cure is sought, namely *hearing* loss.

You might tell me that no audiologist in his right mind could fail to recognise that the sentence test can greatly overestimate the acoustic information a candidate receives via hearing. Yet here is my Consultant surgeon, in charge of one of the largest Programmes in this country:-
*"Open set speech discrimination as a test is used to determine what people can <u>hear</u> of conversational speech without resorting to lip reading. It is believed to be a reasonably sophisticated and subtle technique for assessing <u>how much of the hearing is being used in conversation, rather than the other clues</u> that are available.(*My underlining)
It has been our experience and those of many implant centres around the world that the use of open set speech discrimination both as a tool to assess those most likely to benefit from cochlear implantation and also as a measure of the benefits of cochlear implantation has been widely validated by experience. Ability to have open set discrimination is one of the main determining factors from an audiological point of view as to suitability for an individual for cochlear implant"
This seems to show no awareness of the "clues" provided by what I call supplementary sound interpretation and you call "top-down processing". Giving the most generous possible interpretation, it confuses "hearing" with "understanding" (such an interpretation comes through my gritted teeth because the letter arrived after seven months of my patiently trying to explain why the sentence test was so misleading).

And here is my Consultant audiological scientist and the "Programme Co-ordinator"; *"When assessing whether to proceed with implant surgery, we have to weigh up the probability of whether a patient will in fact be better off following the Implant than they are currently with their Hearing Aids. Our thinking about these probabilities is based on our experience and results with previous patients over many years. Whatever you may think of the speech discrimination test"* (you have a copy of my letter of 20th January to which this was a reply) *", as you currently scored 74% with your Hearing Aids the probability that you would be better off with an Implant is very small indeed. On our current criteria, we do not consider anyone for a Cochlear Implant (except in exceptional circumstances) if they score greater than 30% correct on our speech recognition test. Cochlear Implant surgery involves a complex and difficult operation, and also places at risk any remaining hearing that you have in the implanted ear. We therefore have a responsibility to only operate on those patients when we have a high probability that they are going to be better off following surgery."*

The old saw springs to mind, that it is the duty of a diplomat to fight abroad and lie for his country! Seriously, though, both these apparently respected and honourable people have gone way beyond demonstrating an appropriate degree of loyalty to an established regime imposed by funding authorities and reluctantly lived with. They have either wholly compromised their integrity by mendaciously pretending they believe what they are saying, or there is no pretence - they really do believe that the sentence test can be relied on as they suggest. They really do believe that the performance of one-legged patients selected and fitted with an artificial leg because they were slower than all the rest, can be a reliable guide to what might happen if the faster candidates were given the chance.

Despite the fact that, after eight months correspondence, none of my key points has received a direct reply, I am still reluctant to believe that my correspondents are dishonourable. I am driven to conclude that they may not believe in entering into substantive correspondence with a layman, but strangely it seems far more likely that they are so much in the grip of the sentence test fallacy that it is difficult for them to grasp the points I have made. They are of course people of high intelligence so it ought to be easy for them, but high intelligence can sometimes make it more rather than less difficult to change direction, because a greater weight of intellectual capital has been invested and the inertial force to overcome is more massive.

It does not help that the academic literature presumably reflects the same approach, concentrating on the results of implants measured in terms of sentence recognition. Perhaps that shall only change when people like yourself, who have a firm grasp on how misleading the sentence test is, draw more attention to it. I wonder what your Figure 8 would look like if it plotted the *hearing* of those 119 post-implant patients, and I wonder how large is the gain in *hearing* they each experienced? What was the pre-implant sentence test score, and does the proportionate contribution of "top-down" processing remain the same either side of the operation? I feel that, until this kind of information is compiled, and all the sets of figures are read together, there is little hope of the sentence test fallacy being laid to rest.

So, the present system does amount to an arbitrary form of rationing. For example, my graph at Table "C", for all its over-cautious assumptions, demonstrates that my *hearing* is significantly less than that of candidates for whom an implant would be given without hesitation, yet I fall way outside the guidelines. But I am yet to be convinced that the system is imposed by funding constraints. Rather the arbitrary rationing seems to be operating *unintentionally* and is the result of self-induced intellectual confusion amongst the Cochlear Implant establishment. If you are attending the Conference, you might be able to judge for yourself and I would be fascinated to hear your conclusions.

I very much hope we can meet while you are this side of the water, and I look forward to hearing from you.

With kind regards, Nick Dollar

TELEPHONE CONVERSATION

I have just received a phone call (Sat 11th May 02), which went like this:-

Self: Hello
Caller: Humbledeflumbledeburp (*seemingly mumbled in a woman's voice*)
S [elf]: Sorry, could you repeat that?
C [aller]: Humbledeflumbledeburp
S: I am sorry, come again?
C: Humbledeflumbledeburp
S: I am afraid I am very deaf, could you find another way of saying that?
C: Himbledeflimbledebarp
S: Could you try once more, putting it differently?
C: Hambledeflambledeborp
S: I am sorry, this is getting nowhere. Have we spoken before? (*answer can only be yes or no*)
C: No
S: Are you returning a call I made? (*ditto*)
C: No
S: Is this a business or personal call? (*again only two options*)
C: Business. I am phoning about double glazing.
S: Well thank you but no thank you. Goodbye.

best regards

Reprinted from THE AMERICAN JOURNAL OF OTOLOGY
Vol. 20, No. 4 July 1999
Lippincott Williams & Wilkins Printed in U.S.A.

Jay Rubinstein

The American Journal of Otology
20:445–452 © 1999, The American Journal of Otology, Inc.

Residual Speech Recognition and Cochlear Implant Performance: Effects of Implantation Criteria

J. T. Rubinstein, W. S. Parkinson, R. S. Tyler, and B. J. Gantz

From the Department of Otolaryngology - Head & Neck Surgery, The University of Iowa Hospitals and Clinics, Iowa City, Iowa, U.S.A.

Objective: This study aimed to determine the effects of preoperative speech reception on postoperative speech recognition with a cochlear implant and to develop a statistical index allowing prediction of postoperative speech recognition before implantation. **Study Design:** The study design was a retrospective case review with statistical modeling. **Setting:** The study was conducted at a tertiary referral center with an associated Veteran's Administration hospital. **Patients:** Postlingually deafened adults with and without residual speech reception participated. **Intervention:** Cochlear implantation with Cochlear Corporation CI-22 and CI-24 devices was performed. **Main Outcome Measures:** Monosyllabic word recognition was measured. **Results:** Duration of deafness and preoperative sentence recognition are both significant predictors of word recognition with a cochlear implant and can account for 80% of the variance in word recognition. **Conclusions:** Cochlear implant outcomes are variable but predictable within specific ranges. **Key Words:** Cochlear implant—Speech reception—Residual hearing.
Am J Otol **20:**445–452, 1999.

Evaluation of medical devices such as cochlear implants typically involves comparisons of patient cohorts over time. Ideally, randomized trials compare the performance of one device against another during the same trial. The Iowa cochlear implant clinical trials (1,2) and the Veterans Administration cooperative multicenter trial (3,4) were examples of this type of study design. They clearly showed the superiority of multichannel over single-channel devices and the equivalent performance of the Nucleus WSP and the Ineraid (Cochlear Corp., Englewood, CO, U.S.A.). Since that study, the regulatory process has led to a series of device comparisons that may or may not be valid because selection criteria have changed over time. Specifically, during the U.S. Food and Drug Administration (FDA) clinical trials of a new device, cochlear implant centers focus on testing that device and typically perform fewer surgeries using competing technologies. The result is that outcome data obtained with a new device may not be directly comparable to such data obtained with an older device as the selection criteria for implantation may have changed in the intervening period. Further complicating such studies are measures of outcome assessment, which also may have changed over time and may not be directly comparable.

There have been several unsuccessful attempts to correlate residual hearing before surgery, as measured by pure tones, with cochlear implant performance after surgery (1,2,5–8). The failure to make this correlation has cast some doubt on the hypothesis that spiral ganglion cell survival is a significant determinant of speech reception with a multichannel cochlear implant. Indeed, there is as yet no histopathologic data to support this hypothesis, and the data available suggest the opposite conclusion (9–11). Linthicum has shown that performance of patients with single-channel devices has no correlation with postmortem spiral ganglion cell counts, but speech reception with these devices is poor for all subjects. Preliminary data comparing speech reception during life with spiral ganglion cell counts from a small number of deceased recipients of multichannel devices also show no correlation, but the number of patients still is too small to draw definite conclusions. Duration of deafness has always been the best individual predictor of speech reception with a cochlear implant (2) and also is a strong predictor of spiral ganglion cell populations (12). Although there are insufficient data to examine the relationship between etiology of deafness and speech recognition scores with an implant, the one diagnosis most often associated with poor performance, meningitic deafness, also is highly associated with poor survival of auditory neurons (12).

The following analysis of our recent implant patients will show an association between preoperative sentence recognition in the ear to be implanted and postoperative

Supported in part by research grant DC00242 from the National Institutes of Health/NIDCD; grant RR59 from the General Clinical Research Centers Program, Division of Research Resources, NIH, and the Iowa Lions Foundation (BTG, RST); and by a Biomedical Engineering Research Grant from the Whitaker Foundation, Shannon award DC/OD02948 from NIDCD, and contract NO1-DC-6-2111 from the Neural Prosthesis Program (JTR).

Address correspondence to Dr. J. T. Rubinstein, Department of Otolaryngology, University of Iowa Hospitals & Clinics, 200 Hawkins Drive, Iowa City, IA 52242, U.S.A.

word recognition using the Cochlear Corporation CI-22 and CI-24 devices. We use sentence recognition measures before surgery and word recognition after surgery to avoid "floor" and "ceiling" effects. The analysis verifies prior observations that duration of deafness is the single most important predictor of speech reception with a cochlear implant. With these two preoperative measures, 80% of the variance can be predicted and a preoperative expected performance index has been developed. It is now clear that device evaluations performed under different selection criteria are not directly comparable because of substantial effects of residual speech recognition on performance with a cochlear implant. In addition, expansion of the current criteria may now be appropriate for carefully selected patients using our predictive index.

MATERIALS AND METHODS

The University of Iowa Cochlear Implant Clinical Research Center includes patients implanted at University of Iowa Hospitals and Clinics (UIHC) and the Iowa City Veterans Administration Medical Center (VAMC). Over the past 17 years, House and 3M–Vienna single-channel devices, Vienna 4-channel, Ineraid, Cochlear Corporation CI-22, CI-24, and Advanced Bionics Clarion devices have been implanted. A number of Ineraid subjects have now received Med-El continuous interleaved sampling processor upgrades. Data from multiple implant types will be presented in this article, but detailed analysis will be limited to patients receiving CI-22 devices at the VAMC ("new CI-22 SPEAK") and those receiving CI-24 devices at both VAMC and UIHC; these patients have higher levels of preoperative open-set speech recognition than any prior group we have studied.

Audiometric data

Veterans Administration Medical Center CI-22 subjects (new CI-22 SPEAK) were enrolled in a manufacturer-regulated trial with a testing protocol of 2 weeks, 1 month, 3 months, and 6 months. The CI-24M patients were enrolled in an FDA trial with testing protocol of 2 weeks, 1 month, 3 months, and 4 months. Testing and fitting procedures were identical for both groups. Both groups were programmed in the SPEAK speech coding strategy, although the CI-22 patients used a bipolar stimulation mode while the CI-24 patients were programmed in monopolar mode.

Speech materials were presented before surgery and after surgery in a soundfield at 70 dB sound pressure level from a loudspeaker at 0° azimuth. For each condition, patients were tested with monosyllabic CNC words (CD recordings) with one of ten lists. The same list was never used twice. For CID sentences, one of five lists was chosen for each condition. Subjects adjusted their hearing aid (before surgery) or speech processor (after surgery) to their most comfortable listening level.

To determine candidacy, all patients had preoperative speech recognition tested in the "best-aided condition" as well as having each ear tested alone if binaural aids were used. In the majority of patients, speech recognition in the best-aided condition was the same as the monaural condition for the better ear. In a few subjects, the best-aided condition resulted in CID sentence scores a few percent higher than for the better ear. In all such cases, our analysis is based on speech recognition in the ear that subsequently was implanted. To determine duration of deafness, each patient was asked when he or she stopped being able to use the telephone in the ear to be implanted.

All patients subjected to detailed analysis were postlingually deafened adults with "deep" electrode insertions including all stiffening rings. In all patients, the better hearing, shorter deafened ear was implanted. Except for the five patients excluded below, this cohort includes all VAMC patients with CI-22 implants and every patient at both hospitals who has received a CI-24 device and has at least 3 months' follow-up. Five patients have been specifically excluded:

- One patient was upgraded from a single-channel to a multichannel (CI-24) implant in the same ear. He had received the single-channel device elsewhere 10 years earlier, and the duration of deafness before single-channel implantation is unknown. His speech reception is near the mean for the CI-24 group.
- One patient received radiation therapy for squamous cell carcinoma of the posterior pharyngeal wall discovered during workup for a cochlear implant. The radiation portals included the implanted ear shortly before implantation. This is the poorest performing subject at VAMC.
- Two patients had partial insertions. One of these had chronic otitis media bilaterally and required two procedures to eradicate infection and cholesteatoma in the ear to be implanted. His basal turn required a partial drillout at the time of implantation, and although a complete insertion was obtained at surgery, a partial insertion has been documented on a postoperative Stenver's view. This is our poorest performing subject with a CI-24 device. The second patient had a radiologically normal cochlea, but four stiffening rings could not be inserted. A Stenver's view documents a "270°" insertion, and his performance with the device is substantially less than would otherwise be expected. Blamey (6) has shown that depth of insertion correlates significantly with postoperative speech reception, so only complete insertions without drillouts are included to simplify analysis.
- One patient had postoperative fluctuating attacks of severe roaring tinnitus develop in the implanted ear lasting several hours. There is no vertigo, but her speech reception with the implant fluctuates with the attacks, making definitive measurement of her performance difficult. Her speech reception between attacks is not outstanding ($N = 26\%$), but her exclusion does not alter the final conclusions of this work.

Statistical methods

Linear and nonlinear bivariate analysis was performed on a data set consisting of preoperative CID sentence measures and duration of severe–profound deafness in the ear to be implanted, as well as the postoperative CNC word score at 3, 4, or 6 months. A number of models were tested, but the simplest one yielding high correlations was:

$$pCNC = 53.9 - 0.42 * dur + 0.13 * cid - 0.77 * \left(\frac{dur}{(1+cid)}\right) \quad (1)$$

where pCNC is the predicted CNC word score in percent, *dur* is the duration of deafness in years, and *cid* is the preoperative CID sentence score in percent.

The highest correlation between predicted and actual CNC word scores was produced by a piecewise linear model.

$$\begin{aligned}
0 < dur < 2 \quad & pCNC = 69.1 - 9 * dur - 0.5 * cid \\
2 < dur < 4 \quad & pCNC = -21.2 + 25.4 * dur + 4.3 * cid \\
4 < dur < 26 \quad & pCNC = 51.2 - 1.1 * dur + 0.5 * cid \\
26 < dur < 56 \quad & pCNC = 21.6 - 0.4 * dur + 3.9 * cid
\end{aligned} \quad (2)$$

Although this model produced extremely high correlations, its mathematical discontinuities and instability do not make biologic sense, and it does not allow for extrapolation beyond the modeled data set. A variety of more complex nonlinear models produced somewhat higher correlations than Equation (1) and did allow extrapolation without discontinuities, but their slight improvement did not warrant the additional complexity. Hence, Equation (1) was chosen as our predictive index.

RESULTS

Effects of residual hearing

Figure 1 shows monosyllabic word recognition for all postlingually deafened adults implanted with multichannel devices at the University of Iowa for whom adequate follow-up is available. Means and one standard deviation are plotted for either NU-6 or CNC word lists at 9 months for all devices except the new CI-22 SPEAK data at 6 months and the "CI-24 SPEAK," which includes a mixture of 3- and 4-month data. For the original patients implanted with the CI-22, separate scores are given for F0F1F2, MPEAK, and SPEAK strategies in the same group of subjects. The Med-El continuous interleaved sampling upgrades are on a subset of the patients in the Ineraid group. A number of observations are clear from this summary:

- Mean performance has improved over time.
- Some of this improvement is due to device enhancement—specifically speech processor upgrades—and some is not attributable to devices.
- There is a statistically significant difference between the original CI-22 patients who were upgraded to SPEAK ("original CI-22 SPEAK") and patients more recently implanted with the CI-22 and initially programmed in SPEAK (new CI-22 SPEAK) ($p = 0.021$, t test).
- Speech reception in new CI-22 SPEAK patients is similar to patients using SPEAK on the CI-24.
- A patient-related variable is clearly partially responsible for recent improvements in speech reception.

We became aware of the significant improvement in speech reception in the new CI-22 SPEAK patients over a year ago and suspected that their higher performance was caused by higher levels of preoperative speech reception. Proof of this required the larger subject pool provided by the CI-24 patients. Because performance is similar in the new CI-22 SPEAK and CI-24 SPEAK cohorts, these patients were grouped together for subsequent analysis.

Figure 2 shows monosyllabic word scores as a function of duration of deafness for all patients using the SPEAK strategy. Two separate groups are plotted: "original CI-22" patients is one group and "new CI-22" and CI-24 patients is the other group. Only one patient in the original CI-22 group has any preoperative speech reception because the patients in this group were implanted in the 1980s. Half of the new CI-22 and CI-24 patients have substantial preoperative speech recognition (range, 0%–35%; mean 8% CID; N = 46).

When regression lines are drawn for the two groups, it is clear that duration of deafness is not only an important variable, but it has the same effect on both groups with very similar slopes. Each additional year of profound deafness results in a 0.9% decrease in the expected CNC word recognition for both groups. The improved speech reception of the more recently implanted group is not caused by a different mean duration of deafness as the regression lines clearly show.

The mean difference between the two regression lines is 14%. This represents a patient-related, not device-related, variable responsible for the statistically significant difference between these two groups: the original CI-22 patients

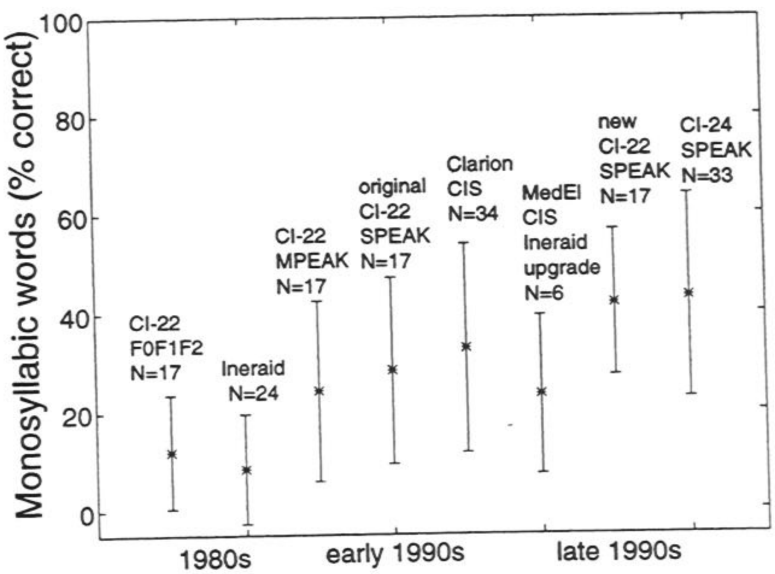


FIG. 1. Monosyllabic word recognition for all adults implanted with multichannel devices at the University of Iowa. Means, one standard deviation, and number of patients (N) are shown. The "F0F1F2," "MPEAK," and "original CI-22 SPEAK" scores are for the same group of patients undergoing speech processor upgrades. There were 25 in this group initially, but only 17 subjects elected to participate in upgrade studies. The Med-El continuous interleaved sampling Ineraid upgrade represents a subset of the Ineraid patients. There is a statistically significant difference in monosyllabic word recognition between the original CI-22 SPEAK patients and the new CI-22 SPEAK patients despite their using the same device. Thus, only device comparisons on the same patients can be made from these data because of the presence of a confounding patient-related variable.

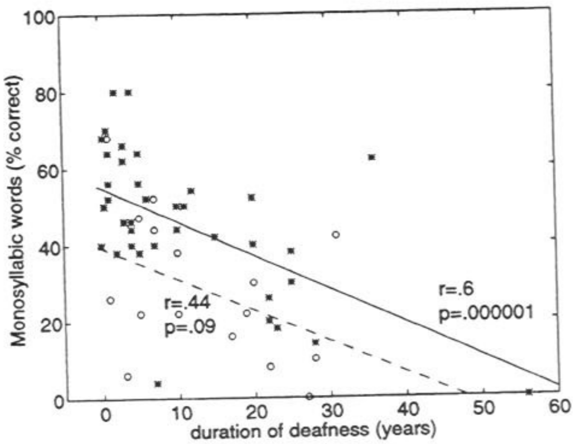


FIG. 2. Monosyllabic word recognition as a function of duration of deafness for patients implanted with CI-22 and CI-24 devices. The asterisks include "new CI-22" and CI-24 patients fitted with a solid regression line. The circles include "original CI-22" patients fitted with a dashed regression line. Note that speech reception in the more recent group is better predicted by duration of deafness than in the original CI-22 group.

implanted in the 1980s and now upgraded to the SPEAK strategy and all patients implanted with the CI-22 and CI-24 in the 1990s and using SPEAK. Current FDA criteria allow for substantially higher preoperative speech reception than was allowed in the 1980s, and therefore the latter group has much greater residual hearing, as measured by CID sentences. The following analysis indicates that this residual speech reception is a major determinant of the different performance in the two groups.

Figure 3 demonstrates the relationship between preoperative CID sentence scores and postoperative CNC word scores in the new CI-22 and CI-24 patients. The postoperative CNC score has a weak, but statistically significant, correlation to the preoperative CID score (= 0.34, $p = 0.019$). The slope of the regression line indicates that for each additional 2% of CID sentence recognition before surgery, an increase of 1% CNC word recognition is expected after surgery. Thus, for a given duration of deafness, if one patient has 40% CID sentences before surgery, that patient would be expected to have 20% higher CNC scores with an implant than a patient with the same duration of deafness who had no open-set speech recognition before surgery. It also is apparent that although those patients with no speech reception before surgery may do extremely well, even small amounts of residual hearing are associated with a higher average performance and lesser variance.

One potential criticism of the comparisons we make is that different tests of monosyllabic word recognition are being used. In our earlier studies, NU-6 words were used for Ineraid, Clarion, and Nucleus F0F1F2, MPEAK, and SPEAK strategies. The later studies of the CI-22 and CI-24 devices used CNC word lists. To eliminate this discrepancy as a possible cause for the effects we report, a group of 19 patients with CI-24 devices were presented both CNC and NU-6 words. The results of this experiment are shown in Figure 4. It is clear that there is no systematic difference between these lists when used in this group of cochlear implant patients.

Another potential criticism of our data is that we have grouped together the new CI-22 patients with those receiving the CI-24. The new CI-22 patients have 6-month data, whereas the CI-24 patients have 3- or 4-month data. Because it is known that performance improves over time (13), our argument that these groups are equivalent can be questioned. Figure 5 shows the CNC word scores for 14 new CI-22 patients at 3 and 6 months.

Nine patients have higher scores and five have lower scores at 6 months. The mean at 6 months is higher than that at 3 months, but the difference is minuscule compared with the variance ($p = 0.38$, t test). This insignificant

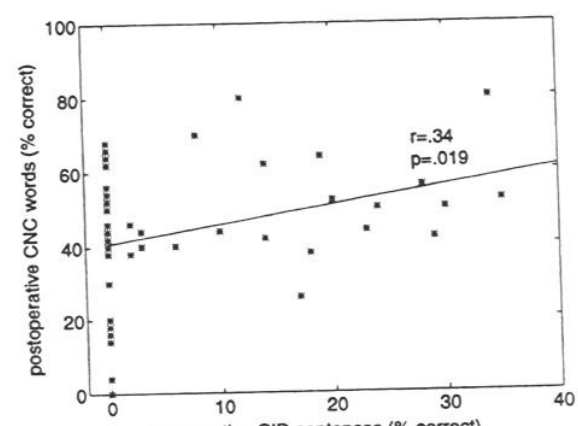


FIG. 3. The CNC score as a function of preoperative CID score for the recent cohort of patients implanted with CI-22 and CI-24 devices running SPEAK. A statistically significant relationship is apparent ($p = 0.019$, = 46).

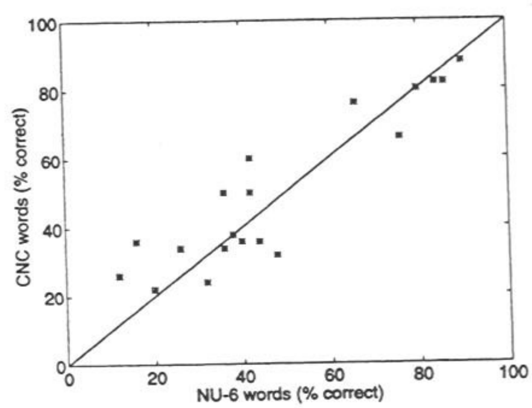


FIG. 4. NU-6 and CNC word lists presented to the same subjects using the CI-24M device. There is no systematic difference in perception of these lists when presented to these cochlear implant patients.

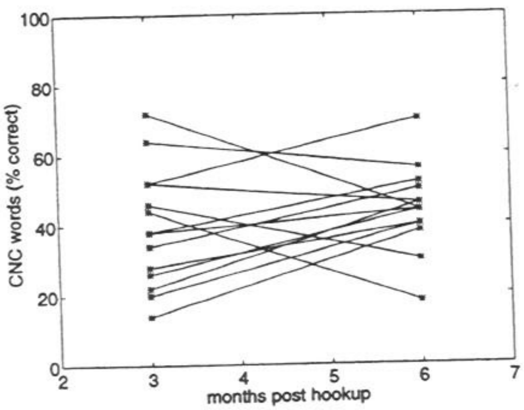


FIG. 5. The CNC word scores for 14 "new CI-22" patients at 3 and 6 months. The mean at 6 months is higher than at 3 months, but the increase does not approach statistical significance.

difference is consistent with the slope of growth of monosyllabic word recognition over time that we have reported previously (13). It remains possible, but speculative, that these subjects who have significant open-set speech recognition before surgery learn to use their implants faster than previous patient cohorts with no open-set speech recognition before implantation. After long-term data are available for these patients, we will be able to answer that question.

A bivariate model for CI performance

We have developed a simple nonlinear model (Equation 1) that can predict outcome with a cochlear implant with a correlation coefficient of 0.77. The more complex, piecewise linear model of Equation (2) can predict outcome with a correlation coefficient of 0.89, but it does not allow extrapolation because of its mathematical structure and its coefficients are not readily interpretable biologically. Other nonlinear models have been developed that have a correlation coefficient of 0.84 and allow extrapolation, but they are far more cumbersome than Equation (1), and the small increase in correlation does not appear worth the added complexity. As we shall see, the correlation of Equation (1) is more than adequate to allow useful predictions of speech reception.

A plot of the actual and predicted CNC score for our patients based on Equation (2) is shown in the left panel of Figure 6. A similar plot based on Equation (1) is shown in the right panel.

Figure 7 shows model calculations from Equation (1) of the effect of duration of deafness on mean word recognition with an implant for two hypothetical patients: one with no speech recognition before surgery and one with 40% CID sentence recognition, the maximum approved by the FDA. There is a substantial difference in the slope of the two lines; the patient with no speech reception before surgery loses postoperative performance at a rate of 1.1% per year of deafness. The patient with 40% CID before surgery loses postoperative performance at less than half that rate, 0.52% per year of deafness. It is clear that having significant residual hearing, as measured by sentence reception, has a substantial effect in ameliorating the deleterious effects of deafness duration.

Figure 8 shows model calculations from Equation (1) of the effect of preoperative sentence recognition on postoperative word recognition with an implant. Two hypothetical cases are calculated: one with 1 year of deafness and one with 30 years. It is clear that for both short- and long-term deafness, preoperative sentence recognition is beneficial to postoperative performance. However, for long-term deafness, even a small amount of residual speech reception dramatically improves performance over that expected with no preoperative sentence recognition. It is interesting that the slope of the functions for the two durations is similar above approximately 10% CID.

Predicting expected outcome and "worst-case" scenarios

The success of the bivariate predictive index in our patient population suggests that we are now in a position to predict

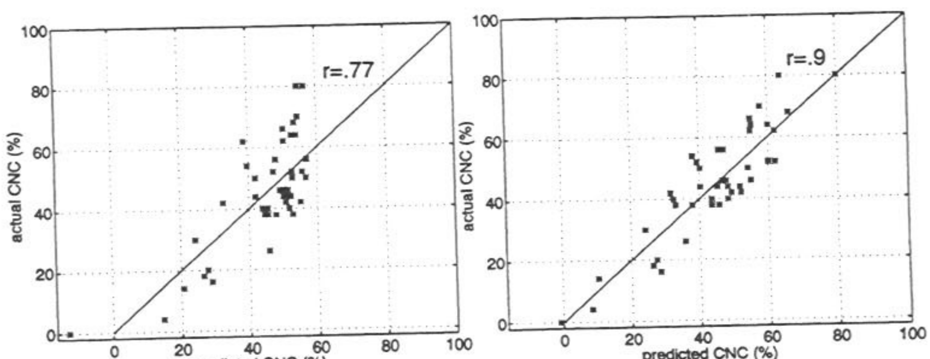


FIG. 6. Predicted and actual CNC scores for our patient population of 46 subjects. The prediction in the left panel is based on the duration of deafness and preoperative CID sentence recognition using Equation 1. The right panel is a similar plot based on Equation 2.

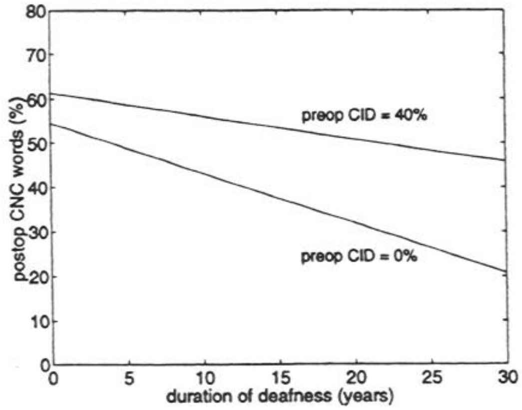


FIG. 7. Predicted postoperative CNC word scores using Equation 1. Two hypothetical patients are calculated, one with no preoperative speech reception (preop CID = 0%) and one with substantial residual hearing (preop CID = 40%).

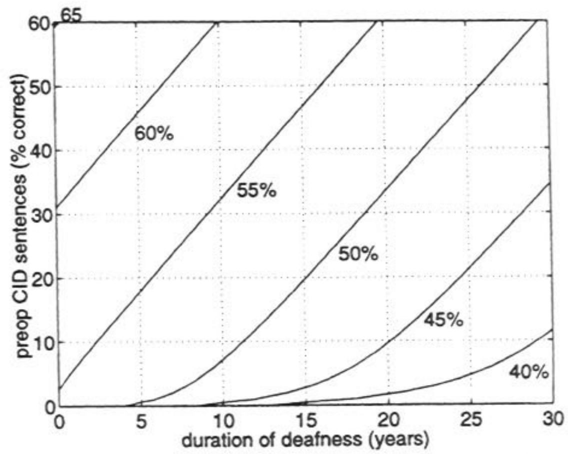


FIG. 9. Contour plot of predicted CNC word scores using Equation 1. The model has been extrapolated to CID values higher than 40% to suggest likely outcomes from implanting patients with hearing better than current criteria allow.

outcomes with a cochlear implant a priori. Figure 9 is a contour plot of Equation (1) showing mean or "expected" CNC word scores as a function of duration of deafness and CID sentence score before surgery. At the intersection of these two values is the mean predicted CNC score. For example, a patient deaf for 10 years with a CID sentence score before surgery of 33% would be expected to have a postoperative CNC word score of approximately 55% 3 to 6 months after hookup. If verified on a larger patient group, this graphical predictive index is a powerful clinical tool for advising potential implant candidates about what is to be expected after surgery.

Given the consistency of the model with a simple biologic interpretation (see below) and given its high predictive power on our patients, we are tempted to assume that the model may apply outside the range of variables with which it was developed. The data suggest that patients above the current FDA criteria of 40% CID sentence recognition may benefit greatly from implantation, and there is no obvious reason to think the model does not apply in the 40% to 60% CID range. However, before subjecting patients with such high levels of residual hearing (e.g., 60% CID) to the risk that the implant may make them worse, a confidence interval must be calculated.

Figure 10 is a contour plot of the 95% confidence interval based on Equation (1). As an example of using this confidence interval, if a patient has been deaf for 15 years and has 10% CID before surgery, there is 95% certainty that

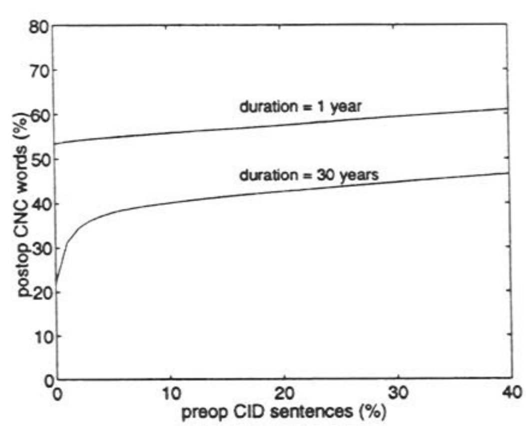


FIG. 8. Predicted postoperative CNC word scores using Equation1. Two hypothetical patients are calculated, one with short-term deafness (duration = 1 year) and one with long-term deafness (duration = 30 years).

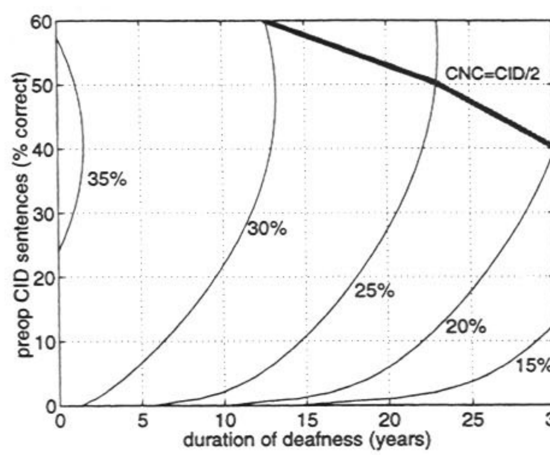


FIG. 10. Contour plot of 95% confidence interval for CNC scores using Equation (1). The model has been extrapolated to CID values higher than 40% to suggest "worst-case" outcomes from implanting patients with hearing better than current criteria allow. The gray bar is a conservative limit beyond which an implant may not be advisable.

this patient will do better than 25% CNC words after surgery. Also shown on this plot is the limit beyond which the predicted postoperative word score is less than half the preoperative sentence score. We include this as a conservative boundary of surgical benefit. At these low levels of speech reception, none of our subjects had CNC word scores that exceeded half of their CID score before surgery.

Extending this analysis to a prospective patient with 40% CID sentences before surgery, currently the maximum residual hearing approved by the FDA, it is clear that such a subject can be implanted out to 30 years with minimal chance that that subject will be worse off after surgery. If we extrapolate to a putative subject with 60% CID before surgery, we can be 95% certain that this subject will have at least 30% CNC words after surgery if his or her deafness is less than 10 years' duration. The mean performance for such subjects would be expected to be more than 60% CNC words (Fig. 9).

DISCUSSION

Although we have not accounted for all variability with cochlear implantation, we have made great progress in identifying two significant, yet simply obtained, preoperative variables that predict outcome. We previously have identified other preoperative predictors (2), which would account for even more of the variability but which require more elaborate preoperative testing. Duration of deafness and preoperative sentence recognition are standard measures easily obtained on all implant candidates at most centers.

A simple biologic explanation for these clinical correlations is apparent. The data imply that patients with short durations of deafness are likely to obtain high levels of speech reception with a cochlear implant regardless of their preoperative speech recognition. Because most severe–profound hearing loss is of cochlear origin, a large population of spiral ganglion neurons is expected for short durations of deafness (12). For those patients deafened for many years, if no speech reception is present, few inner hair cells are functioning, and without their trophic support, secondary degeneration of the spiral ganglion is unimpeded. If substantial speech reception persists, the residual inner hair cells protect the spiral ganglion from the effects of long durations of deafness. One could also postulate that residual speech reception maintains the viability of the central auditory pathways, but such trophic effects are not as clearly defined as they are at the periphery. Our clinical correlations cannot differentiate between these two hypotheses; they only suggest that residual speech reception has a trophic effect on performance with a cochlear implant.

The variance data also are consistent with this "speech trophism" hypothesis for the new CI-22 and CI-24 subjects. The standard deviation of the CNC scores for the patients with no preoperative speech reception ($CID = 0$) is 19.2 ($N = 24$). For the patients with some speech reception ($CID > 0$), the standard deviation is 13.6 ($N = 22$). The difference in the sample variances is significant ($p = 0.05$, F-distribution). Because the difference is not likely caused by chance, one possible interpretation is that the difference is caused by spiral ganglion cell survival. For patients with no speech reception before surgery, spiral ganglion cell counts could, in theory, vary between 0 and 30,000. Those patients with some speech reception before surgery ($CID > 0$) must have enough inner hair cells and spiral ganglion cells to support their rudimentary speech recognition. Thus, if spiral ganglion cell populations are important for cochlear implant performance, a smaller variance would be expected for the group with residual speech reception.

This hypothesis also can explain why there is poor correlation between residual pure-tone thresholds and performance with a cochlear implant. In the somatosensory system, activation of only one peripheral neuron is adequate to achieve psychophysical threshold (14). Identification of an object held in the hand, however, presumably requires a more substantial number of peripheral receptors and neurons. It seems safe to assume that the auditory system is similar and that achievement of psychophysical threshold for a pure tone requires fewer spiral ganglion cells than perception of open-set speech. This is presumably the reason why loss of speech discrimination is a typical early sign of retrocochlear disease, so-called auditory neuropathy (15).

Our findings do not directly bear on the choice of ear for implantation, but they do lend strong support to our current practice of implanting the better hearing, shorter deafened ear. They also suggest that in selected cases, criteria for implantation could be expanded with well-defined risks. The average monosyllabic word score is now 45% and for patients with residual speech reception is more than 50%. Our predictive index indicates increasing performance with greater levels of preoperative speech reception. The expected performance and confidence intervals for durations of less than 10 years suggest that implantation criteria could be expanded to 60% CID in selected patients with a minimal chance of making any patient worse, assuming complete (25-mm) insertions can be reliably obtained by the surgeon. There is a significant likelihood of performance levels in such subjects substantially exceeding those currently obtained. We hope to test this hypothesis in a small group of appropriately informed subjects in the near future.

REFERENCES

1. Gantz BJ, Tyler RS, Knutson JF, et al. Evaluation of five different cochlear implant designs: audiologic assessment and predictors of performance. *Laryngoscope* 1988;98:1100–6.
2. Gantz BJ, Woodworth GG, Abbas PJ, Knutson JF, Tyler RS. Multivariate predictors of audiological success with multichannel cochlear implants. *Ann Otol Rhinol Laryngol* 1993;102:909–16.
3. Cohen NL, Waltzman SB, Fisher SG. The Department of Veterans Affairs Cochlear Implant Study Group. A prospective randomized study of cochlear implants. *New Engl J Med* 1993;328:233–7.
4. Cohen NL, Waltzman SB. Multicenter comparative study of cochlear implants: final reports of the Department of Veterans Affairs Cooperative Studies Program. *Ann Otol Rhinol Laryngol* 1995;104(Suppl 165).
5. Blamey PJ, Pyman BC, Gordon M, et al. Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients. *Ann Otol Rhinol Laryngol* 1992;101:342–8.
6. Blamey PJ. Pre- & Post-Operative Factors Affecting Speech Perception in Adult Cochlear Implant Users. 1997 Asilomar

Conference in Implantable Auditory Prostheses. Pacific Grove, CA.
7. Waltzman SB, Fisher SG, Niparico T, et al. Predictors of postoperative performance with cochlear implants. *Ann Otol Rhinol Laryngol* 1995;104(Suppl 165):15–8.
8. Battmer RD, Gupta SP, Allum-Mecklenburg DJ, et al. Factors influencing cochlear implant perceptual performance in 132 adults. *Ann Otol Rhinol Laryngol* 1995;104(Suppl 166):185–7.
9. Blamey P. Are spiral ganglion cell numbers important for speech perception with a cochlear implant? *Am J Otol* 1997;8(6 Suppl):S11–2.
10. Linthicum FH, Otto SR. Functional histopathology of four 22-electrode cochlear implant temporal bones. *ARO Abstracts* 1997:82.
11. Fayad J, Linthicum FH Jr, Otto SR, et al. Cochlear implants: histopathologic findings related to performance in 16 human temporal bones. *Ann Otol Rhinol Laryngol* 1991;100:807–11.
12. Nadol JB, Jr, Young Y-S, Glynn RJ. Survival of spiral ganglion cells in profound sensorineural hearing loss: implications for cochlear implantation. *Ann Otol Rhinol Laryngol* 1989;98:411–6.
13. Tyler RS, Parkinson AJ, Woodworth GG, et al. Performance over time of adult patients using the Ineraid and Nucleus cochlear implants. *J Acoust Soc Am* 1997;102:508–22.
14. Ruch TC. Somatic sensation: receptors and their axons. In: Ruch TC, Patton HD, eds. *Physiology and biophysics I: The brain and neural function*. Philadelphia: Saunders, 1979.
15. Starr A. Clinical findings in patients with auditory neuropathy. Conference on Auditory Neuropathy. Lake Arrowhead, CA, 1998.