CHAPTER

1

Presidential Address: "How Do You Know?"

STEPHEN J. HAINES, M.D.

Perhaps never in the history of medicine have we been so challenged to answer such a simple sounding question as: "How do you know?" From the daily problems of patient care ("How do you know the right way to do a cervical discectomy?" "How do you know the right treatment for sciatica?") to matters of broader patient-care policy ("How do you know the value of reoperation for glioblastoma?") and new interventions ("How do you know the value of radiosurgery for trigeminal neuralgia?"), from seemingly mundane financial affairs ("How do you know the cost of running your practice?") to weighty considerations of broad scope ("How do you know the value of neurosurgery to the health-care system?"), we are asked to document our beliefs with more objectivity, reliability, and reproducibility than ever before. So today, and for the rest of this meeting, and for the rest of your practicing lives, I ask you to take a journey with me to try to understand how it is that we do know that what we do is the right or best thing to do.

LEARNING TO COUNT

We know that much of the authority of practitioners of healing arts in the past relied on appeals to divine or mystical power. Some of the earliest great observers of human anatomy and physiology had their pronouncements enshrined as dogma that controlled the practice of medicine for centuries. The idea that careful tabulation of the results of treatment might lead to better understanding of the best way to treat patients is relatively new.

Pierre Charles Alexandre Louis promoted a method including:

- careful observation and description of clinical details
- systematic record keeping
- rigorous analysis of multiple cases

Presidential Address delivered at the 44th Annual Meeting of the Congress of Neurological Surgeons, Montreal, Quebec, Canada, September 30, 1996.

- cautious generalizations based solely upon observed facts
- verification through autopsy (1, 3).

In 1825 he published an influential study tabulating the results of treatment of tuberculosis with bleeding and finding no evidence of a beneficial effect (8). This and subsequent studies in typhoid fever refuted the dogmatic reliance on bleeding that dominated medicine at that time.

In 1837 there was a great debate in the French Academy of Sciences between Louis and the defender of the status quo, Monsieur Double (4, 9). Essentially, Double claimed that the individual variation between patients exceeded any tendency toward group similarities, observing that "... numerical and statistical calculations, open to many sources of fallacy, [are] in no degree applicable to therapeutics."

Louis' claim was that "... a therapeutic agent cannot be employed with any discrimination or probability of success in a given case unless its general efficacy in analogous cases has been previously ascertained .... Therefore I conceive that without the aid of statistics nothing like real medical science is possible." In essence, Louis concluded that the effect of the treatment must exceed the variation in effectiveness between patients if the treatment is to be advocated for a specific condition. Many American physicians came to Paris to learn from Louis. Tabulation gradually became a standard approach for determining the value of a treatment.

PARALLEL DEVELOPMENT OF NEUROSURGERY AND BIOSTATISTICS

Neurosurgery was not very far advanced at that time, and one would have expected the new field to take advantage of the opportunity to use the most advanced newly developed tools of evaluation to secure its place in the medical armamentarium. Neurosurgery and biostatistics developed in parallel during the last 150 years, and it is an interesting exercise to apply a developmental assessment to neurosurgery's use of evaluative technology.

Table 1 lists the major advances in neurosurgery and biostatistics during the past 150 years as I see them. It was a bit difficult to construct this table for statistics, but the list comes from an historical work by Gehan and Lemak (5) and from discussions with several practicing statisticians.

Where is neurosurgery on the evaluation time line? I asked this question one way in 1979 (6). Reviewing all clinical articles published in the *Journal of Neurosurgery* up to that point, most of them used the most rudimentary evaluation techniques: tables developed by Louis in the 1820s and statistics from the early 1900s. Only 18 of 4865 at-

TABLE 1.
Advances in Neurosurgery and Biostatistics

Time	(Bio)statistics	(Neuro)surgery
Forever	Appeal to the gods	Appeal to the gods
1st millennium AD	Appeal to dogma	Appeal to dogma
1800–1850	Counting/tabulation	Anesthesia
1850–1900	Probability theory, normal distribution, standard deviation	Antisepsis
1900s	Student's t-test, χ^2 test	Sphygmomanometry
1910s	Maximum likelihood methods	Electrocautery
1920s	Experimental design and ANOVA	Ventriculography
1930s	Multivariate analysis	Angiography
1940s	RCT	Myelography
1950s	Odds ratio/relative risk	Steroids
1960s	Life tables	Microsurgery
1970s	Logistic regression	CT scan
1980s	Meta-analysis	Magnetic resonance imaging (MRI) scan
1990s	Exploratory data analysis	Radiosurgery

tempted to use controls, 10 of these with randomization and just 1 with blinding techniques. In the 1970s we were learning to use computerized tomography (CT) scans in neurosurgery, and we were evaluating them with techniques from the 19th century.

If we look at the major advance of the 1990s, radiosurgery, we find that, despite a smattering of more sophisticated techniques, most of the evaluation is done with ancient and outmoded technology. It is as if we were operating with ether, monopolar cautery, and without steroids or the microscope.

There has been a large paradigm shift in the understanding of how to find the truth that is buried in our observations and interpretations of it: a shift from faith and dogma to repeatable observation and hypothesis testing. Technological advances have dominated changes in neurosurgery over the same period of time. We are just beginning to understand the fallacies of observation and to apply modern techniques to evaluation.

PAVING STONES ON THE ROAD OF GOOD INTENTIONS

Why all the fuss, some will say. Are not advances in surgery self-evident? No—and sometimes our desire to help our patients gets in the way of our understanding how best to do it. Why do we have so much trouble making good evaluations of what we do? Mostly because we want to succeed—to have our patients do well because we have done our jobs well. Those good intentions may lead us to overlook some of the reality checks that we encounter along the way.

I would classify these good intentions that may lead to a distorted view of success as:

- the desire to do good and well
- the *surgeon's* selective memory
- the patient's kindness
- and the old saw, "only time will tell."

Desire to Do Good and Well

Let me tell the extracranial-intracranial (EC-IC) bypass story with a perspective different from what you have probably heard before. Believing that carotid endarterectomy prevented ischemic stroke by restoring normal blood flow, and knowing that some patients suffering ischemic stroke had stenotic lesions beyond the reach of endarterectomy, wise and technically sophisticated neurosurgeons reasoned that restoring intracranial blood flow by bypassing the stenosis should prevent stroke. My first mentor in neurosurgery, Pete Donaghy, and his first microsurgical fellow, Gazi Yasargil, developed and simultaneously performed the first EC-IC bypasses. To many, it was self-evident that this was the correct solution to the problem and when the technical success of the procedure was demonstrated, it was declared the standard of therapy.

Unfortunately, it was another beautiful theory destroyed by a few dirty facts. When applied to a population of patients who accurately represented the population undergoing the operation in actual practice and after the results were carefully and objectively observed, recorded, and analyzed, it was found that the operation had no net benefit over nonoperative therapy.

A few still think the benefit is self-evident. I have heard a distinguished neurosurgeon say that he cannot see one of these patients living with the threat of stroke from inaccessible stenosis and send him home without doing *something* for him. This is a laudable emotion but does not justify an ineffective operation. He says "but I know that the operation works," and I say: "Show me—how do you know?"

The Surgeon's Selective Memory

Another problem with our traditional way of evaluating surgical progress is the (selective) surgical memory. One of my colleagues is fond of telling us of the importance of surgical memory: the ability to remember great results and forget bad ones. He says it is essential to the practice of neurosurgery, for anyone who dwells on poor outcomes soon becomes ineffective. He is as fond of telling us that he has never seen a complication of a certain procedure as we are of reminding him of the last time it happened.

In many ways he is right: in day-to-day practice we must focus on the positive. However, what is good for daily practice may not be good for policy making. Our daily optimism must be based on sound observation or else we will make daily errors. How often do we find that our personal estimate of the rate, say, of postendarterectomy stroke or shunt infection is noticeably lower than what we find when our last 100 cases are carefully reviewed? It is very difficult to accurately estimate low-frequency events from memory and when we do we frequently fool ourselves.

The Patient's Kindness

Our patients also desperately want us to succeed. Many are reluctant to tell us about our failures. I have done a large number of microvascular decompressions for trigeminal neuralgia and hemifacial spasm with what I think are excellent results. I try to be open to patient concerns and encourage comments about less than optimal outcomes.

The patients have been grateful and supportive of the operations. When we hired a nurse-clinician to help care for these patients, we found what seemed to be an unusually high incidence of persistent postoperative headache. It turned out that the patients were so glad to be relieved of their primary symptom (and probably, so happy to be alive after brain surgery) that they simply didn't want to bother me with their headaches.

We've essentially solved the headache problem by doing cranioplasties at the end of the procedures, but we wouldn't have even known about the problem without getting someone other than the surgeon involved with the postoperative follow-up. My point is that what seems self-evident—the success or failure of neurosurgery—isn't, and that this is not just an academic concept but applies to daily practice.

"Only Time Will Tell"

One of the most common approaches to evaluation is simply to keep doing a new operation until the weight of opinion determines its usefulness or lack thereof—"Only time will tell." The statement suggests that there is no other (or at least no better) way of doing the evaluation. This is manifestly not true. Table 2 summarizes data comparing the evaluation of neurosurgical therapy in five content areas using either randomized clinical trials or nonrandomized techniques.

In each content area, the value of the treatment under study was not clear after years of nonrandomized evaluation took place. When the number of patients evaluated in inconclusive nonrandomized studies is

TABLE 2.
Controlled versus Uncontrolled Studies in Neurosurgery

	Uncontrolled		RCT		Ratio
Content Area	No. of Studies	*N*	No. of Studies	*N*	Uncontrolled/RCT
Antifibrinolytic therapy	21	3,398	8	479	7.1
Chemonucleolysis	>20	>20,000	3	234	>85.5
Antibiotic prophylaxis	15	13,787	5	2,713	6.3
EC-IC bypass	23	2,662	1	1,377	1.9
Carotid endarterectomy	51	17,484	3	1,626	10.8

compared to the number required to obtain a definitive evaluation in RCTs, the results are stunning. The ratio of inconclusive nonrandomized patients to conclusive randomized patients ranges from 2 to 1 to greater than 85 to 1. Clearly, advanced clinical research techniques can find a result much faster and more efficiently than trusting to the passage of time.

Indeed, the passage of time may obscure the true value of a procedure. A useless procedure may seem beneficial because related therapy has improved. A useful procedure may not seem so because of a change in diagnostic criteria, decline in severity of disease over time, and so on. There are many ways better than the passage of time to evaluate new procedures, such as case-control studies, cohort studies, and clinical trials.

FIFTY-SEVEN WAYS TO FOOL YOURSELF

David Sackett, the first North American physician since Osler to be appointed to the Professorship of Medicine at Oxford, has catalogued many ways of fooling one's self at every stage of studying a clinical phenomenon (11). Table 3 emphasizes that there are a multitude of traps in the process.

TOOLS OF KNOWING

The tools available to us fall into four categories:

- logic
- training
- experience
- experimentation or science

Logic is often the first tool used. Faced with a new situation, for which we have no experience or training, we argue from basic principles to

TABLE 3.
Fifty-Seven Ways to Fool Yourself

- Literature review
  - Rhetoric
  - "All's well"
  - One-sided reference
  - Positive results
  - "Hot stuff"
- Sampling
  - Popularity
  - Centripetal
  - Referral filter
  - Diagnostic access
  - Diagnostic suspicion
  - Unmasking
  - Mimicry
  - Previous opinion
  - Wrong sample size
  - Admission rate
  - Prevalence/incidence
  - Diagnostic vogue
  - Diagnostic purity
  - Procedure selection
  - Missing data
  - Chronology
  - Starting time
  - Unacceptability
  - Migratory
  - Membership
  - Nonrespondent
  - Volunteer
- Intervention
  - Contamination
  - Withdrawal
  - Compliance
  - Therapeutic personality (placebo)
  - Bogus control
- Outcome measurement
  - Insensitive measure
  - Rumination
  - End digit
  - Apprehension
  - Unacceptability
  - Obsequiousness
  - Expectation
  - Substitution
  - Family information
  - Exposure suspicion
  - Recall
  - Attention
  - Instrument
- Analysis
  - Post-hoc significance
  - Data dredging
  - Scale degradation
  - Tidying-up
  - Repeated peeks
- Interpretation
  - Mistaken identity
  - Cognitive dissonance
  - Magnitude
  - Significance
  - Correlation
  - Underexhaustion

arrive at a plan of action. This usually requires that we assign different values or weights to the facts available. Often what seems "logical" to one person seems illogical to another because the basic principles used or the weights assigned are different.

For example, when a managed health-care plan sets up shop in a community which has not had one before, it seems logical to some physicians to fight, refuse to participate, and organize against the plan, whereas to others it seems logical to adapt, understand the rules by which they operate, and participate while maintaining as much advantage as possible. Logic is obviously a blunt tool, sometimes the only one available, but is very fallible and leads to inconsistent, frequently ineffective, solutions.

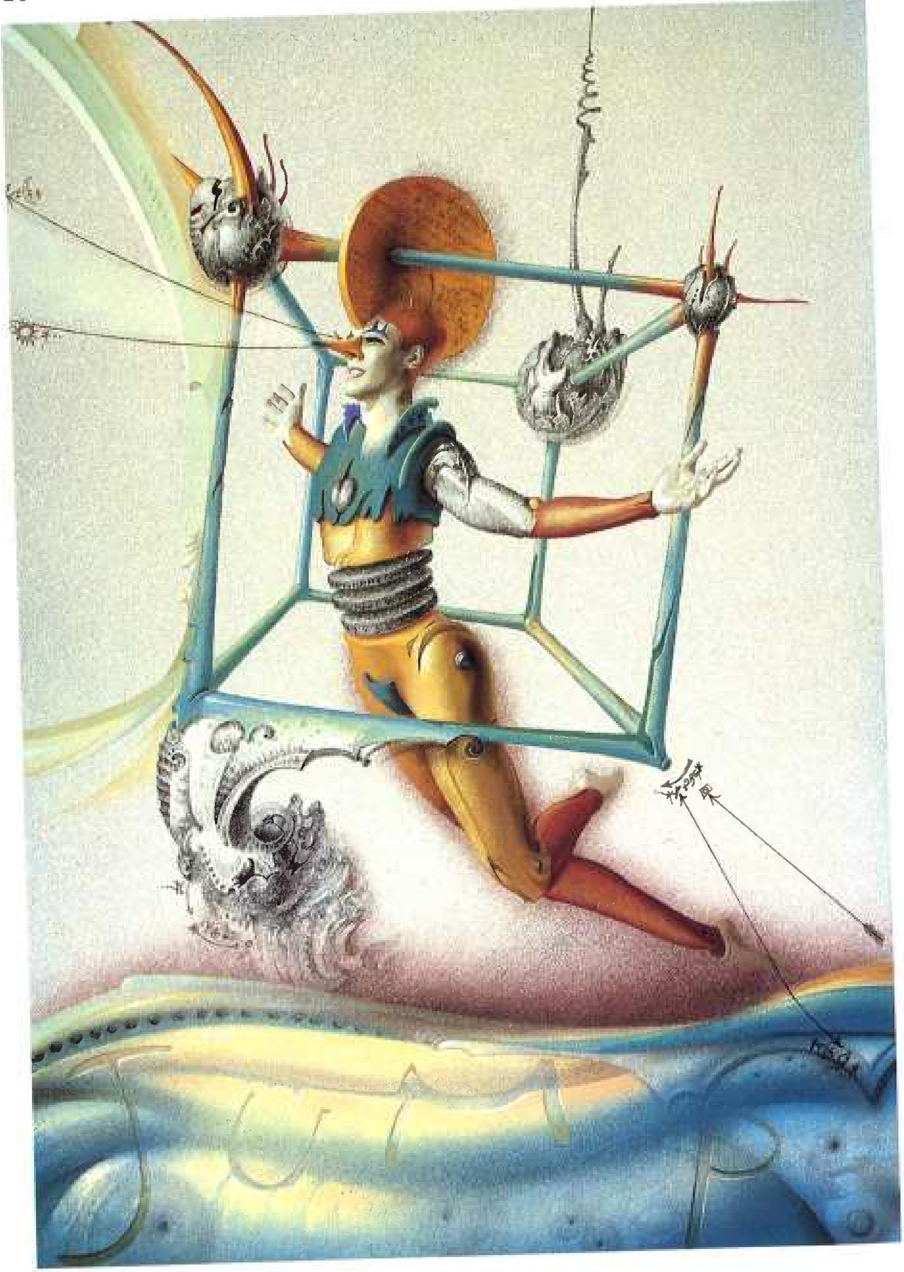
"Rock Dove" 1982 oil on wood 110 × 90 cm © 1996 Michael Parkes/Steltman Galleries New York

The second tool is training. When one wishes to become expert in dealing with a set of situations and there already exists a number of experts in the field, that expertise can be acquired most efficiently by learning from those experts. This is essentially organized experience—learning from others' mistakes.

"Juggler" 1981 oil on wood 110 × 90 cm © 1996 Michael Parkes/Steltman Galleries New York.

I have had the very good fortune to learn from a number of outstanding neurosurgeons:

- my father, Gerald Haines, from whom I have learned more about the attitude, energy, and discipline needed for the practice of neurosurgery than anyone else could teach me
- Pete Donaghy, the most truly humble pioneer neurosurgery has ever known
- Peter Jannetta, whose ability to create an atmosphere of learning, inquiry, and innovation is unparalleled
- Shelley Chou, a man of patient intelligence, who conquers every problem to which he applies his considerable intellectual skills and inspires respect from all who work with him
- Roberto Heros, who has an unusual ability to guide the hands of those with whom he is operating



Jump
Gil Bruvel

The Spiral of Wisdom
Gil Bruvel

Without the benefit of their experience, distilled through the training process, I could not be a neurosurgeon. We all have a similar story.

Training has the positive value of efficiency. But it is not enough by itself. It may also perpetuate dogma, superstitious behavior, and unrecognized error. What is learned is only as good as the training, experience, and analytic skills of the mentor, filtered by his or her teaching ability.

Experience refines our knowledge, training, and logical deductions and inferences. Positive experience reinforces what we have done, promoting successful behavior but also preserving unnecessary behavior, superstition. For example: an experienced neurosurgeon, call him Harvey, has an extensive and successful experience in lumbar disc surgery. He has always given 10 mg of dexamethasone at the end of the case because he was trained to do so, and he attributes his unusually short average length of stay (for his hospital) of 36 hours to this practice. He associates with an equally experienced neurosurgeon, we will call him Walter, who has never given dexamethasone for routine lumbar discectomy (and, in fact, thinks it's a crazy idea). After 6 months in practice together it is clear that they have identically good results, length of stay, and other indicators of quality.

This suggests to me that the dexamethasone is unnecessary, but perhaps Harvey is so heavy handed that he needs to use the dexamethasone in order to have results as good as his partner. Neither knows if the other is right (although we can assume that each thinks the other wrong). The point, however, is that just because things are going well does not mean that everything being done is necessary to obtain such good results. This wasn't a big problem when resources seemed unlimited, but when the unnecessary expenditure of some resources may result in the unavailability of others, such questions need to be asked. Neither Harvey nor Walter has asked the question, "How do you know?"

Negative experience extinguishes the behavior we blame for the bad result. Unfortunately, unless we analyze things correctly, we may eliminate the wrong behavior. Therefore, experience, while a good teacher, is best at: refining a base of knowledge and skill acquired from more rigorous scientific investigation and adapting general knowledge and skills to our individual abilities. Experience allows us to *artfully interpret* the knowledge and skill we acquire from others to make it most effective when applied by our hands to an individual patient. So each of these tools has an important role, but their true value comes in inverse order to that which we usually employ.

Too often we treat patients based on a logical inference from something passed on to us in training, modifying it as we gain experience (at the expense of our earlier patients), and resort to clinical science only when there is enough controversy to create a political need for scientific results. The bedrock of our actions as a scientifically based profession should be a core of knowledge acquired through rigorous scientific observation and experimentation in the clinical arena. This generates basic principles of care generalizable to many patients. We acquire most of this knowledge through training, through which it is passed on to us with the benefit of avoiding the errors of the past. We refine the application of this general knowledge through our own experience and, when faced with new situations in which we must act before acquiring knowledge, we use logic to develop an action plan from known principles.

VALUE OF THE QUESTION

So what is the value of asking "How do you know?" I hope that it is becoming clearer that it is the best way to have a sound basis for clinical action and make those actions as effective and efficient as possible. With such a high stake in successful outcomes and so many ways to fool ourselves, a little informed skepticism is necessary. Asking the question regularly can improve your daily practice of neurosurgery.

But we are not the only ones in need of some informed skepticism. As resistant as we may be to asking ourselves "How do you know,?" I sense no reluctance on the part of neurosurgeons in asking insurers, HMOs, the Health Care Financing Administration, or the Clintons, "How do you know:

the value of neurosurgical services?"
the need for a certain procedure?"
the right way to deliver health care?"

These are good questions; we should ask them forcefully and demand data to support their decisions that is as good as the data they demand from us. The Congress and the AANS along with the JCSNS are asking these questions regularly and as forcefully as possible through the Joint Washington Committee for Neurosurgery.

There are other good questions, such as: "How do you know that cognitive medicine (thinking about illness and health) is more valuable than procedural medicine (doing something about it)?" or "How do you know that the tort system is the best way to assure quality in the practice of medicine and compensation for losses caused by medical error?" or "How do you know whether you are reimbursing us too much for the expense of running our practices?"

THE SCIENTIFIC LIFE

Where does this leave us? We have chosen a caring and helping profession in which we artfully interpret a body of scientific knowledge for the welfare of our patients. This is what I would call the *scientific life.* The scientific life is one in which we seek to act with predictable outcomes based on careful, valid, and reproducible observations.

However, the science *in* our lives does not eliminate the art *from* our lives. Artful interpretation and application of scientifically derived principles to individual situations is what we do, but we seek to act from a solid base of observed fact. We need to emulate the leopard boldly leaping from the firm base of the mountain into the unknown rather than the juggler balanced in some mysterious way in thin air, sometimes providing a mystically enchanting performance, but too often falling to his death (10). Science and art are inextricably linked as Douglas Hofstadter in his seminal work, *Goedel, Escher and Bach* (7), so elegantly pointed out, or as Gil Bruvel shows in his paintings (2).

Science is indifferent to dogma and authority. Science serves us well as patients, as a profession, and as a society, but only if we are ever skeptical and ask the question, "How do you know?"

It has been a great privilege and responsibility to serve you this year. I have tried to do so on the basis of the best information available, to improve the quality of that information, to artfully interpret and apply it based on my training and experience, and to ask at each critical juncture, "How do you know?" I hope that you all feel the same sense of privilege and responsibility in continuing to seek excellence in providing neurosurgical care to your patients and ask yourselves and those around you, "How do you know?"

REFERENCES

1. Bollet AJ: Pierre Louis: The numerical method and the foundation of quantitative medicine. **Am J Med Sci** 266:92–101, 1973.
2. Bruvel G: The Reality of a Dreamer. Fort Lauderdale, FL, Apropros Art Gallery, 1992, plates "Jump" and "The Spiral of Wisdom."
3. Cassedy JH: *American Medical and Statistical Thinking 1800–1860.* Cambridge, MA, Harvard University Press, 1984, pp 60–64.
4. Double: The inapplicability of statistics to the practice of medicine. **Am J Med Sci** 21:247, 1837.
5. Gehan EA, Lemak NA: *Statistics in Medical Research.* New York, Plenum Medical Book Company, 1994.
6. Haines SJ: Randomized clinical trials in the evaluation of surgical innovation. **J Neurosurg** 51:5–11, 1979.
7. Hofstadter DR: *Goedel, Escher, Bach. An Eternal Golden Braid.* New York, Basic Books, 1980.
8. Louis PCA: Anatomic, pathologic, and therapeutic research on the disease known by the name of gastroenteritis putrid fever, adynamic atoxic typhoid, etc. **Am J Med Sci** 4:403, 1829.

9. Louis PCA: Louis on the application of statistics to medicine. **Am J Med Sci** 21:525, 1837.
10. Parkes M: *Michael Parkes*. Amsterdam, Steltman Editions, 1993, Plates 54 and 95.
11. Sackett DL: Bias in analytic research. **J Chronic Dis** 32:51–63, 1979.