Chapter 1

Clinical evidence

Zoher Ghogawala MD FACS, Director 1; Chairman, Department of Neurosurgery 2; Associate Professor of Neurosurgery 3
Bob S Carter MD PhD, Professor and Chief of Neurosurgery 4
Fred G Barker II MD, Associate Professor of Neurosurgery 5

1 Wallace Clinical Trials Center, Greenwich, Connecticut, USA
2 Lahey Clinic Medical Center, Burlington, Massachusetts, USA
3 Tufts University School of Medicine, Boston, Massachusetts, USA
4 University of California, San Diego, California, USA
5 Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, USA

Introduction

The practice of neurosurgery has traditionally been taught by experienced surgeons who share their wisdom through an apprenticeship over a period of a several-year residency. While this strategy has worked well to teach surgical techniques, there has been a relative lack of instruction regarding the implementation of clinical evidence. What is clinical evidence? It is the evidentiary support that follows from the application of the scientific method to study a clinical question. Clinical evidence goes beyond one surgeon's accumulated experience and derives power from the scientific application of rigorous methods to determine the effectiveness of a treatment that applies to all surgeons and their patients.

There are several reasons why evidence-based decision-making is not inculcated into neurosurgical training; the most prominent of which is the lack of class I evidence or well-designed randomized controlled trials (RCTs) to guide our specialty [1, 2]. Second is a general lack of understanding regarding how to interpret clinical evidence of lower levels than an RCT. Making matters more complex, of the few RCTs that do reach completion, many do not provide a definitive answer to the research question or, even if they do, may fail to convince practitioners to alter their established patterns. Today, there is a growing need to supplement our education with a systematic understanding of the principles of evidence-based practice. In this chapter, we outline the key elements of clinical evidence. These elements are used throughout this textbook to aid our understanding of specific studies that might ultimately guide neurosurgical practice.

Outcomes

Well-designed clinical studies require the application of valid outcome measures. Many types of outcome instruments are available. The validation of these tools is the focus of much clinical research. The

utility of outcome instruments is that they provide comprehensive data regarding the success of any clinical intervention that goes well beyond a surgeon's clinical assessment of 'good,' 'fair,' or 'poor' results. Payers, hospitals, physicians, and patients all have different perspectives when considering outcomes. In this book, we focus on the perspectives of the patient and clinician. We present five areas of outcome analyses: clinical outcomes, surrogate outcomes (e.g., imaging outcomes), functional outcomes, quality of life outcomes, and cost-effectiveness outcomes. Cost outcomes are becoming particularly important when evaluating surgical options. Clinicians will need to become more familiar with this terminology going forward.

Clinical outcomes

Clinical outcomes measure objective clinical events associated with a given disease entity. Discrete clinical events (e.g., death, stroke, specific neurological outcomes) are, by definition, clinically relevant and can be very useful in measuring the clinical efficacy of neurosurgical interventions when the natural history of a disease is associated with major morbidity or mortality. Mortality (survival) as a clinical outcome is generally used in malignant diseases where life expectancy is limited. For example, Kaplan-Meier survival curve [3] analysis has been used in malignant brain tumor trials where death is expected within months to years of the initial diagnosis [4]. For carotid occlusive disease, on the other hand, where disease-specific mortality is rare in the short term, the efficacy of carotid endarterectomy has been demonstrated largely by comparing stroke rates for different treatment strategies. NASCET (North American Symptomatic Endarterectomy Trial), for example, demonstrated that for symptomatic patients with ipsilateral carotid stenosis of 70% or greater, carotid surgery reduced the risk of ipsilateral stroke from 26% (best medical management) to 9% (carotid endarterectomy plus best medical management) at 2 years [5]. This randomized trial (comprising 2,226 patients) convincingly demonstrated the value of surgery over best medical management and has become the basis for the standard of care.

Surrogate outcomes

Surrogate outcomes are proxies for true clinical outcomes. The most commonly used surrogates in neurosurgery are radiographic outcomes (e.g., achievement of lumbar fusion as a correlate for relief of back pain, percent coil obliteration as a surrogate for preventing aneurysmal re-hemorrhage). Surrogates can also be non-radiographic, such as recording intracranial pressure monitoring data in studies of head trauma.

Imaging outcomes are often used as surrogate measures for clinical outcomes to obtain an early prediction of the eventual clinical outcome. In some situations, radiographic data might permit early intervention to modify the clinical course before symptoms are manifest. MRI evidence of tumor recurrence, for example, is an outcome measure used in many studies of malignant brain tumors where treatment effects might be modest but significant [6]. Radiographic studies for fusion, stability, or deformity assessment are often used as surrogate measures for eventual clinical outcome in spinal studies [7]. Rigorous studies using imaging outcomes often employ independent blinded reviews, with well-defined characteristics for each outcome measure.

Because imaging outcomes are often available earlier than clinical outcomes (e.g., MRI evidence of tumor recurrence usually precedes death), they avoid the confounding effect of effective salvage therapies (other than the therapy being tested in the trial). A major drawback of many imaging outcomes, however, is their failure to correlate with outcomes that are clinically significant. For example, many studies have found that radiographic spinal fusion has little or no correlation with any meaningful clinical outcome, such as pain relief [8, 9].

Functional outcomes

Functional outcomes are focused evaluations of a patient's state of function that do not attempt to summarize a patient's entire state of health. Examples of functional scores include pain perception scales (e.g., Visual Analogue Scale [10]), disability scales (e.g., Oswestry Disability Index) [11], 'return to work' status,

and the Karnofsky Performance Scale [12]. Disease-specific functional outcome scales aim to quantify the degree of disease-specific impairment at baseline or to measure the change in impairment status after an intervention. When discussing changes in disease-specific outcomes scales, it is critical to address the degree of change that is clinically meaningful – that is, the minimum clinically important difference. For example, many degenerative spine trials consider a difference of 10 points on the Oswestry Disability Index to be clinically significant [13, 14].

General health-related quality of life outcomes

General health-related quality of life (HRQoL) measures are commonly used in neurosurgical studies to measure patients' overall view of their entire state of health. (The difference between 'quality of life' and HRQoL is that 'quality of life' embraces many factors outside the medical realm, such as satisfaction with work and personal relationships.) The Short Form (SF)-36, -12, and -8 instruments are all questionnaires completed by patients that are used to score HRQoL. In the SF-36, a commonly reported HRQoL instrument used in many studies, raw scores for eight categories (physical functioning, role [physical], bodily pain, general health, vitality, social functioning, role [emotional], and mental health) are linearly transformed using scoring algorithms to yield scores from 0 to 100. A physical component summary (PCS) or mental component summary is then generated using population-adjusted norms to generate normalized scores with a mean±standard deviation of 50±10. Many different medical disciplines have used the SF-36 PCS to assess the impact of an intervention on HRQoL. For example, a successful hip replacement is associated with a 9.6 point improvement in the SF-36 PCS score [15]. The minimal clinically important difference for the SF-36 PCS score has been reported by some to be at least 5 points [13, 16].

Cost-effectiveness outcomes

Cost-effectiveness analysis requires the measurement of clinical effectiveness with a preference-based HRQoL outcome score (that has been scaled to equal 0 for death and 1 for perfect health) for the purpose of calculating quality-adjusted life-years (QALYs) as a measure of the difference between two treatments. One common example of a preference-based HRQoL instrument is the EuroQol-Five Dimensions (EQ-5D) questionnaire [17].

Cost data from an intervention can be combined with QALY outcome data to generate cost-effectiveness data. In essence, the life expectancy after one treatment is multiplied by the expected HRQoL after that treatment; the same is done for the other treatment, and the difference is the QALY difference between the treatments. The cost difference between the treatments is then divided by the QALY difference to yield the cost per QALY gained. The cost of hemodialysis (arbitrarily, $50,000 per QALY gained) is often used in the USA as a benchmark for determining whether a new intervention is cost-effective [18, 19]. The specific cost-effectiveness threshold typically varies from society to society; for example, the National Institute for Health and Clinical Excellence, which advises the National Health Service in the UK, recommends a cost-effectiveness threshold of £20,000-30,000 per QALY gained for approving new therapies – often controversially [20].

The specific method used to calculate costs can greatly affect the cost-effectiveness ratio. In SPORT (Spine Patient Outcomes Research Trial), the cost-effectiveness of surgery relative to non-operative treatment for lumbar disc disease was $69,403 per QALY using overall adult surgery costs (all payers), and $34,355 per QALY using Medicare population-specific surgery costs [21]. It is common to estimate hospital costs by reporting hospital charges, which are readily available. Real hospital costs can be estimated using coding systems such as the Healthcare Common Procedure Coding System. In addition, Medicare reimbursement rates for specific hospital billing codes (International Classification of Diseases, 9th Revision [ICD-9] and Diagnosis-Related Group [DRG]) are readily available and are also used to estimate the costs of health care [22].