September 30, 2019

Nina F. Schor, MD, PhD, Deputy Director
Office of the Director
National Institute of Neurological Disorders and Stroke
National Institutes of Health
Bethesda, MD 20892

> **SUBJECT:** **NINDS strategic plan planning process input from national organized neurological surgical organizations**

Dear Dr. Schor:

On behalf of the American Academy of Neurological Surgeons (AAcNS), American Association of Neurological Surgeons (AANS), American Board of Neurological Surgery (ABNS), Congress of Neurological Surgeons (CNS) and Society of Neurological Surgeons (SNS), representing organized neurosurgery in the United States, we are pleased to respond to your request for input on the National Institute of Neurological Disorders and Stroke (NINDS) strategic plan planning process. Below are our collective thoughts and recommendations related to the following points, as defined in the NINDS request for information.

I.   **Opportunities for, and challenges facing, progress in neuroscience research and neurological health.**

Neurosurgeon researchers provide a unique opportunity to advance biomedical understanding and the development of new therapies for nervous system disorders. Neurosurgeons can provide insights into neurobiological-clinical correlations across the continuum of neurologic disease. Neurosurgeon-driven understanding will come from the assessment of patients that can be directly linked to operating room observations and direct response to monitoring and therapy. Traditionally, neurosurgeons form and propel essential collaborations with investigators across scientific fields that have culminated in a deeper understanding of normal and disease pathobiology. These findings have led to new effective treatment paradigms (medical and surgical) for previously untreatable or ineffectively treated nervous system disorders. Neurosurgical access to the nervous system can provide direct and potent insights into nervous system anatomy, biology and physiology. Moreover, neurosurgeons are essential to the use of new surgical therapeutic technology for treatment.

By effectively exploiting the distinctive opportunities that neurosurgeon-scientists use to impact neuroscience research, there are a number of NINDS-related mechanisms that could be employed that would have an immediate and transformative impact in lessening the burden of neurologic disease.

- First, expand the support of successful in-training and early career support, including the R25, K12, K08 and K23, that have multiplied independent neurosurgeon investigators and neuroscience research impact.

- Second, device and technologic development are essential in neurosurgery. Grants targeted at surgical technologies are, therefore, critical.

- Third, growing support for observational natural history and treatment studies will be transformative to optimizing understanding and management neurologic disease.

- Fourth, neurosurgeons can more effectively translate nervous system therapies/technologies through more robust support.

- Finally, the creation of a pool of expert clinical trialists who use randomized trials and prospective observational database methodologies will be crucial to turning basic and translational discoveries into effective clinical therapies.

II.   Emerging scientific or organizational trends, advances, techniques, and perspectives NINDS should account for in this planning process.

There are emerging scientific trends, advances and techniques in neurosurgery that can significantly and positivity impact neurologic disease in a short period of time (3 to 5 years). These will require collaborative efforts across multiple disciplines. Support to expand and/or mature the following nervous system advances will have an impact:

- **Vascular device development.** Endovascular devices are rapidly evolving and improving. These emerging devices will provide more effective and safer treatments for ischemic stroke, hemorrhagic stroke and aneurysms through design and linked therapeutics.

- **Responsive nervous system stimulation.** Closed loop (sensing/responsive) deep brain stimulation devices are emerging for clinical use and may provide more effective personalized stimulation paradigms for the management of neurodegenerative disease, epilepsy, addiction and/or psychologic disorders.

- **Gene therapy and gene-editing technology.** Direct genetic manipulation of diseased neural circuitry can have a potent and near-term impact on a wide variety of neurologic disorders that are currently not treatable or ineffectively treated. It is a rich area for collaboration across multiple neuroscience research areas.

- **Robotics.** Cranial and spinal robotics are emerging as opportunities for the more effective and efficient surgical treatment of cranial and spinal disorders. Advances in these devices should improve quality and consistency across a large domain of neurosurgical procedures.

- **Artificial Intelligence and machine learning.** Data analytics are being used in the neurosurgery. These provide an opportunity to enhance understanding and optimal management of neurologic disease, while potentially reducing the number of patients needed to draw conclusions. These analytics permit accurate and personalized diagnoses, as well as management/outcomes across related disease populations. The use of artificial intelligence processes to analyze large, prospective observational databases will be a powerful tool to refine the diagnosis and management of neurosurgical diseases.

- **Wearables and digital phenotyping.** By collecting real-life/activity information, wearables and digital phenotyping provide an opportunity to obtain valuable objective data, understand disease state and/or therapeutic impact. They can also better inform artificial intelligence and machine learning paradigms in the clinical neurosciences.

- **Minimal access technologies**. These technologies include high- and low-frequency ultrasound, stereotactic radiosurgery and laser interstitial thermal therapy for the treatment of neurologic disorders. These technologies continue to expand. Understanding their impact on the natural history of disease and comparing their effectiveness to other management options will determine their therapeutic value.

- **Targeted pharmacologic therapies.** New directed pharmacologic treatments developed through basic mechanistic and pathway studies need to be translated into clinically effective

therapies. This will require a group of well-trained clinical trialists and support for IND-enabling studies.

- **Modulation of the neuro-immune system.** Emerging data indicate that neuroimmune modulation is becoming more critical in the treatment of neurologic disorders. Support for collaborations across research areas that were not previously linked is important.

- **Single-cell genomics.** Single-cell genomics will be essential to define neurologic disorders at the cellular level. Application of these findings is essential to broader disease understanding and the development of target therapeutics.

- **Advanced imaging technology.** Advanced imaging technologies will drive biologic understanding, improvement in non-surgical/surgical treatment and therapeutic assessment in many neurologic disorders. Evolving understanding of disease can be optimized by non-invasive imaging techniques. This will also drive therapeutic discovery and monitoring.

- **Biomarkers for disease.** Developing imaging, blood, cerebrospinal and/or other biomarkers will be essential for early diagnosis (premorbid), tracking disease progression and monitoring of therapeutic effectiveness. Support in this area could rapidly change the scope of disease, diagnostic paradigms and management.

- **Defining the natural history of neurologic disease**. Defining the natural history of a variety of neurologic disorders, including neoplasia, vascular disease, trauma, spinal disorders and neurodegenerative diseases, will have a profound impact on indications for treatment, expected outcomes and comparing the effectiveness of various disease-modifying therapies.

III. **Recommendations for steps, actions, activities, and opportunities that will enable NINDS to make rapid progress toward achieving each goal.**

Resident, fellow and early career research training support

The R25, K12, K08 and K23 have supported the rapid expansion of neurosurgeon independent investigators over the last four years. The vitality of these programs is critical to growing critical neuroscience discoveries.

- Expand support of in-training (residency and fellowship) and early career (0 to 5 years after training completion) neurosurgeon-scientist awards.

- Define new and innovative mechanisms with national neurosurgical organizations for the funding research training that are responsive to changes in the neurosurgical training environment (e.g., changes in residency requirements are research training and financial complexities of an early career neurosurgeon).

- Creating a robust pool of young neurosurgeon scientists will lead to innovative physiologic, diagnostic and therapeutic strategies across all of the neurosciences.

Define funding streams that support early and late device and technologic development

This is an essential area of need in neurosurgical research. Because the specialty is by definition device and technology-driven, defining innovative strategies that support the early and late stages of device and technology development will be critical to moving device-related technologies forward. Neurosurgeons can effectively translate invasive nervous system therapies/technologies and increasing support of these tools can be foundational. This could be supported in several ways:

- Expansion of intramural-extramural opportunities to drive technology-related trials forward where appropriate.

- Grow opportunities for NIH, industry and academic partnerships to move device/technologies forward.

- Define support for regulatory training and IND/IDE-enabling studies to provide a pathway to the clinic for promising new technologies.

Support for prospective observational studies for comparative effectiveness research and determining natural history

To optimize management and develop new therapies for neurologic disease patients, it will be critical to define the natural history of the diseases we treat.

- Support prospective, observational database studies that include sophisticated biostatistical techniques such as propensity score matching and predictive analytics. These studies will complement randomized, controlled trials to increase participation and knowledge generation in the clinical neurosciences.

- Properly designed prospective observational databases allow meaningful comparative effectiveness research to be done without randomization and include more representative patients and surgeons than is often the case with randomized trials.

Developing a pool of expert neurosurgeon-scientists who are clinical trialists

Funding the creation of a pool of expert neurosurgeon clinical trialists, skilled in designing and conducting randomized trials and/or prospective observational database comparative effectiveness studies, will be essential to developing effective therapies from basic and translational science discoveries.

- Expansion of intramural-extramural opportunities to drive translational research and early clinical trials.

- Define support for regulatory training and investigation new device (IND)/investigational device exemption (IDE)-enabling studies to provide a pathway to the clinic for promising new technologies.

- Grow and support opportunities for NINDS, industry and academic partnerships to move device/technologies forward.

IV. **For recommendations you make, please consider indicating appropriate objective success criteria, including quantitative and qualitative benchmarks and milestones for gauging progress in the corresponding domain.**

For the recommendations defined above, we provide objective success criteria, as defined below.

Criteria for success of resident, fellow and early career research training support

- Criteria of success for the R25 and K12 awards would be defined by:
    - Conversion rate to K08, K23 or other R01/DP; and/or
    - Conversion rate to other equivalent federal funding.

- Early-career awards (within first five years of completion of training) including the K08 and K23 would be defined by:
    - Successful R01 or DP-awards; and/or
    - Conversion rate to other equivalent federal funding.

Criteria for success of support for early and late device and technologic development

- Increase in the number of NINDS, industry and academic trials supporting early or late device development;

- Development of new targeted training programs with an increase in neurosurgeons involved in IND/IDE-enabling studies; and/or

- Increase in the number of NINDS supported IND/IDE-enabling studies for new technologies.

Criteria for success of support for observational or natural history studies

- Increase in the number of NINDS supported prospective observational database or studies for comparative effectiveness and natural history research;

- Published definition for specific disease natural history; and/or

- Increase in prospective, observational trials that use sophisticated biostatistical methods such as propensity score matching that will complement randomized control trials and increase participation and knowledge generation in clinical neurosciences.

Criteria for success of support for developing a pool of neurosurgeon-scientist clinical trialists with expertise in randomized, controlled trial and prospective, observational database methodologies

- Increase in the number of intramural-extramural opportunities related to translational research and early clinical trials;

- Increase in the number of investigators supported for regulatory training and IND/IDE-enabling studies; and/or

- Increase in the number of NINDS, industry and academic partnerships to move early clinical trials forward.

V. Successes, shortcomings, and impacts of existing NINDS policies, practices, partnerships, strategies, or activities.

Neurosurgery has seen success from the NINDS research training awards. These have led to increases in early career, independent National Institutes of Health (NIH) and other federal funding at a significantly higher rate than those individuals who have not been supported via these awards. This has driven an increase in the number of independent neurosurgeon-scientists.

Shortcomings include the requirements for departmental cost-sharing to support surgeon-scientist, clinical service requirements for residents who elect to spend research time during training and the difficulty associated with translating neuroscience device-related technologies via NIH funding mechanisms.

VI. Stakeholders and experts NINDS should consult in the process of strategic planning and strategy implementation.

We would recommend the following representatives across the various neurological surgery stakeholders:

- One representative (neurosurgeon) that is a member of the NINDS Advisory Council;

- One representative for the Neurosurgery Research and Education Fund (NREF);

- One representative from the Executive Committee of the Neurosurgeon Research Career Development Program (NRCDP);

- The president/chair (or designated representative) from each of the following national neurosurgical organizations: the American Academy of Neurological Surgery, Society of Neurological Surgeons, American Association of Neurological Surgeons, American Board of Neurological Surgeons, Congress of Neurological Surgeons and AANS/CNS Washington Committee,

VII.   High-priority objectives that you do not see reflected among the four stated strategic planning goals.

We do not see clinical trial conduct and design reflected in the strategic planning goals. It will be important that a pool of surgeon-scientists is educated in the design and conduct of clinical trials in the neurosciences, including the acquisition of scientific correlates and information from the clinical trials. Similarly, we do not see funding prospective observational studies that will be valuable for comparative effectiveness research and more generalizable to "real-life" clinical practice than traditional randomized control trials.

Thank you for the opportunity to provide input into this important process. If you have any questions or need additional information, please do not hesitate to contact us.

Sincerely,

E. Antonio Chiocca, MD, PhD, President
American Academy of Neurological Surgeons

Christopher I. Shaffrey, MD, President
American Association of Neurological Surgeons

Linda M. Liau, MD, PhD, Chair
American Board of Neurological Surgeons

Ganesh Rao, MD, President
Congress of Neurological Surgeons

Karin M. Muraszko, MD, President
Society of Neurological Surgeons

Attachment: "Launching Effectiveness Research to Guide Practice in Neurosurgery: A National Institute Neurological Disorders and Stroke Workshop Report"

Staff Contact:
Katie O. Orrico, Director
AANS/CNS Washington Office
25 Massachusetts Avenue, NW, Suite 610
Washington, DC 20001
Direct: 202-446-2024
Fax: 202-628-5264
Email: korrico@neurosurgery.org

Patricia Walicke, MD, PhD*
Aviva Abosch, MD‡
Anthony Asher, MD§
Fred G. Barker, II, MD¶
Zoher Ghogawala, MD||
Robert Harbaugh, MD#
Lara Jehi, MD**
John Kestle, MD‡‡
Walter Koroshetz, MD*
Roderick Little, PhD§§
Donald Rubin, PhD¶¶
Alex Valadka, MD||||
Stephen Wisniewski, PhD##
E. Antonio Chiocca, MD, PhD***
for the Workshop Participants*

*National Institute Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland; ‡Department of Neurosurgery, University of Colorado-Anschutz Medical Campus, Aurora, Colorado; §Department of Neurosurgery and Neuroscience Institute, Carolinas Healthcare System, Charlotte, North Carolina; ¶Department of Neurosurgery, Massachusetts General Hospital, Boston, Massachusetts; ||Alan L. and Jacqueline B. Stuart Spine Research Center, Department of Neurosurgery, Lahey Hospital and Medical Center, Burlington, Massachusetts; Tufts University School of Medicine, Boston, Massachusetts; #Department of Neurosurgery, Penn State University, Milton S. Hershey Medical Center, Hershey, Pennsylvania; **Epilepsy Center, Cleveland Clinic, Cleveland, Ohio; ‡‡Department of Neurosurgery, University of Utah, Salt Lake City, Utah; §§Biostatistics, University of Michigan, Ann Arbor, Michigan; ¶¶Department of Statistics, Harvard University, Cambridge, Massachusetts; ||||Department of Neurosurgery, Virginia Commonwealth University, Richmond, Virginia; ##Department of Epidemiology, University of Pittsburg, Pittsburgh, Pennsylvania; ***Department of Neurosurgery, Brigham and Women's Hospital, Center for Neuro-Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts

Correspondence:
E. Antonio Chiocca, MD, PhD,
Department of Neurosurgery,
Brigham and Women's Hospital,
75 Francis St,
Boston, MA 02493.
E-mail: EAChiocca@partners.org

Launching Effectiveness Research to Guide Practice in Neurosurgery: A National Institute Neurological Disorders and Stroke Workshop Report

This workshop addressed challenges of clinical research in neurosurgery. Randomized controlled clinical trials (RCTs) have high internal validity, but often insufficiently generalize to real-world practice. Observational studies are inclusive but often lack sufficient rigor. The workshop considered possible solutions, such as (1) statistical methods for demonstrating causality using observational data; (2) characteristics required of a registry supporting effectiveness research; (3) trial designs combining advantages of observational studies and RCTs; and (4) equipoise, an identified challenge for RCTs. In the future, advances in information technology potentially could lead to creation of a massive database where clinical data from all neurosurgeons are integrated and analyzed, ending the separation of clinical research and practice and leading to a new "science of practice."

KEY WORDS Clinical trials, Effectiveness research, Randomized controlled trial, Registry, Equipoise

O ver the last decade, 2 themes have emerged as major drivers for the future of health care: comparative effectiveness research and precision medicine. Both have relied heavily on data derived from clinical trials, particularly the randomized controlled clinical trial (RCT). Technological developments such as electronic health records (EHRs) will vastly increase the accumulation and availability of data related to outcomes of care, potentially transforming clinical research and reducing the reliance on the RCT. In fact, EHRs and prospective registries are now thought by some to provide information that could complement or replace information derived from RCTs.

Whether RCT or registry is used, the analyses must be statistically rigorous and scientifically sound to best guide the evolution of practice so that 21st century neurosurgery will be based on reliable, scientific data, and analyses. The information provided by either RCT or registry may also lead to more effective use of medical resources and reduce expense of care.

The Launching Effectiveness Research to Guide Practice in Neurosurgery Workshop was held on February 5, 2015 in Bethesda, Maryland, to discuss clinical research in neurosurgery. The attendees included experts in neurosurgery, neurology, statistics, effectiveness research, and bioethics, along with representatives from federal agencies including National Institutes of Health (NIH), Food and Drug Administration (FDA), Center for Medicare and Medicaid Services (CMS), and Agency for Healthcare Research and Quality (AHRQ). The Research Committee for the Society for Neurological Surgery and National Institute Neurological Disorders and Stroke (NINDS) leadership worked together to ensure broad participation for a variety of stakeholders. For example, for the neurosurgery/neurology participation, selection of participants was based on current leadership in national neurosurgical societies, principal investigatorship on NIH, or other type of grants

related to neurosurgical or neurological research in clinical trials, and participation in clinical trials related to neurosurgery. Similarly, representatives for federal agencies were selected based on their expertise and/or leadership in the matter under study. This report is a summary of the meeting agreed upon by the authors.

BACKGROUND: CLINICAL TRIALS AND RESEARCH IN NEUROSURGERY

The RCT is widely accepted as the gold standard in clinical research, but published RCTs are infrequent in neurosurgery. Kiehna et al[1] found approximately 10 clinical trials published each year in the 5 leading neurosurgical journals. A more recent analysis of the English language literature from 2000 through 2014 identified 61 neurosurgical RCTs, with an average annual publication rate of approximately 5 neurosurgical RCTs.[2] The paucity of RCTs reflects, in part, performance barriers in the surgical setting, and frustration with limited generalizability of surgical RCTs. Clinical research in neurosurgery is more likely to employ designs like observational cohorts or retrospective case series.

Multiple factors have shaped clinical trials design, particularly including the differences between efficacy and effectiveness.[3-5] Efficacy is the observed effect under ideal circumstances, and is the upper limit of attainable benefit. Effectiveness trials, also called pragmatic trials, are meant to determine the degree of beneficial effect in "real-world" clinical settings. In any clinical research design, there is a trade-off between studying efficacy and effectiveness. A study designed to assess efficacy (with limited entry criteria for patients and surgeons, tightly controlled interventions, and a focused, validated outcome measure) will be at risk of limited generalizability. Conversely, relaxing these parameters may make the results more generalizable, but risks compromising internal validity. The prototypical efficacy (or "explanatory") study is the double-masked, randomized, parallel arm controlled study often required to support approval of a drug or device by the FDA. Effectiveness research has tended to employ observational designs, though these have evolved from simple description to more complex approaches (see below) which may allow valid drawing of causal inferences.

RCTs and descriptive observational studies each have different strengths and limitations. The RCT has high internal validity, randomization being the best method for managing sources of bias and confounding. However, the RCT often has low external validity—conclusions from many RCTs do not generalize to daily neurosurgical practice, because they fail to incorporate real-world variation in patient mix, surgical skill and training, and other aspects of clinical care like nursing skill and the availability of critical care facilities and physical therapy. They also may fail to be relevant as newer technologies emerge or as proficiency with techniques continue to evolve over time.

In contrast, registry-based observational trials can have high external validity, provided that the registry captures complete, sequential data on the population or a representative sample of the population. However, the internal validity of observational trials is weak, because estimates of treatment effects are subject to bias from unobserved confounding factors. When results of observational trials and RCTs are compared for the same disorder and interventions, there is a clear tendency for the former to overestimate benefit.[1,6–8] In some cases, the bias may lead to a false-positive result, for example, in the early studies of glioblastoma brachytherapy, which were marred by confounding. In this example, the apparent benefit from brachytherapy treatment after glioblastoma resection was spurious, arising from the favorable prognostic factors that defined eligibility for the treatment.[9-11] More often, observational trials and RCTs do concur on whether or not there is benefit from a treatment[6,8] ($r = 0.75$)[7].

There are meaningful operational and logistical differences between the prototypical extremes of RCT and registry observational study. On a per-patient basis, the implementation of RCTs tends to be more labor intensive for both the central organization and the clinical sites, and therefore more expensive. In principle, an observational database can potentially be used repeatedly for multiple studies, further improving cost efficiency. In practice, observational studies may be compromised by missing data particularly on patient-reported outcomes.[12] Although RCTs can be large, involving tens of thousands of patients, observational studies ultimately have the advantage of larger scale, with examples such as the FDA-supported Mini-Sentinel database, which captures treatment and safety information on millions of patients.[13] Logistics tend to limit the duration of RCTs to a decade or less, whereas observational data collection can be open ended. In addition to their different scientific and technical properties, operational differences may also guide the choice between an RCT and an observational study in specific situations.

Because randomized treatment assignment confers high internal validity, the RCT remains the gold standard for demonstrating treatment efficacy within the confines of the study. However, evidence is needed for effectiveness of procedures and therapies in diverse neurosurgical practice. Restrictive RCTs cannot address the effectiveness of each step and instrument in neurological procedures conducted by diverse surgical teams in varied practice settings, with heterogeneous patients. Such questions require the more flexible and cost-efficient approaches of observational trials. However, these trials need rigorous statistical design and analysis, in order to provide reliable results to guide neurosurgical practice.

STATISTICAL REQUIREMENTS FOR RIGOROUS DEMONSTRATION OF EFFECTIVENESS FROM OBSERVATIONAL DATA

Effectiveness implies a causal relationship between an intervention and an outcome. Under the widely accepted framework sometimes referred to as the Rubin Causal Model (RCM),[14,15]

the causal effect is a "comparison of the outcome that would be observed with the intervention and without the intervention, both measured at the same point in time". The causal effect is inherently impossible to measure, because only the outcome under the assigned treatment can be determined for any individual patient. RCTs randomize treatment assignment to generate cohorts of comparable individuals who receive different interventions. Deriving statistically valid causal inference from observational data requires reproducing as closely as possible the characteristics of an RCT. A proxy must be found for the randomization method: a means of data management or segmentation is applied to generate comparable cohorts for comparison. The propensity score, an estimate of the probability of assignment to treatment or control given the multiple variables potentially available to the clinical decision maker, serves this purpose. Estimating a valid propensity score requires considerable reflection on the types of data that influence the decision makers, and might include inherent patient characteristics, prognostic factors, training or experience of the surgeon, factors related to the hospital, economic factors like insurance, and patient and family preference. It is not unusual for more than 20 variables to contribute to the propensity score.[14,15]

An RCT differs from a well-designed prospective observational study in one design issue—the use of randomization to allocate patients to treatment groups. Randomization reduces the risk that treatment effects, within the trial, will be distorted by known or unknown confounders. Treatment effects within the trial can then be determined by directly comparing outcomes. In a prospective observational study, treatment selection is influenced by factors that may differ among groups, so we must account for these differences when determining treatment effect.

Propensity score matching is a way to more accurately determine treatment effects in nonrandomized trials by controlling the existence of known confounding factors. Comparison of outcomes using treated and control subjects who are as similar as possible on a wide range of potential confounders allows us to perform a pseudo-randomized study and draw causal inferences using a nonrandomized, prospective observational study design. The extent to which propensity score matching can eliminate bias from confounders depends on the completeness and quality of the control variables on which the propensity score is computed and the matching performed. Propensity score matching cannot eliminate bias that may be introduced by unknown confounders.

The propensity score is defined as the conditional probability of receiving a treatment given pretreatment characteristics. To design a prospective observational database study that maximizes the effect of propensity score matching, we need to ask: (1) What randomized experiment do we want to model, (2) Who are the decision makers for treatment assignment, (3) What are the key covariates used to assign treatment, (4) Can we measure the key covariates well, (5) What clinically meaningful outcomes will we measure, and (6) What sample sizes will be needed?

Therefore, characteristics of RCTs that should be duplicated to achieve reliable results from observational data include the following:

- Prospective specification of outcomes and analytic methods without resort to the actual outcome data, even if this is already available
- Prospective estimation of sample size
- Outcome measures that are clearly defined and captured in the study data.
- Completeness of the data set for outcomes, that is, limiting missing data.
- The ability to measure and record all of the important covariates that influence treatment assignment.

The improved power and accuracy of causal inferences after generation of comparable cohorts using propensity scores were exemplified at the workshop using the International Study of Unruptured Intracranial Aneurysms (ISUIA) observational data set for unruptured intracranial aneurysms.[17] In ISUIA, the risk of subarachnoid hemorrhage from an unruptured aneurysm was dependent on several factors including aneurysm location, aneurysm size, and aneurysm morphological features, and the management outcome after surgery was dependent on patient age, aneurysm location, and aneurysm size. In ISUIA, patients were not randomized to a procedure or to conservative management. They were selected for surgery, endovascular therapy, or conservative management based on the clinician's recommendation for each patient. Therefore, any comparison of surgical and conservative management needed to take into account potential differences in the cohorts. In a preliminary analysis using propensity scores in the aggregate data, surgery was statistically significantly superior to conservative care for both hemorrhage and overall outcome at both 5- and 10-year time points ($P \leq .001$).[17]

A number of modern analytic methods seek to uncover causal relationships from observational data, including instrumental variables,[18] marginal structural models,[19] and inverse probability weighting.[20] All of these methods of analysis have utility in some circumstances, but all require recourse to the outcome data and thus are subject to change with different outcome variables. Therefore, ultimately "design trumps analysis for objective causal inference".[15] At the workshop, 2 highly experienced trialists (Robert Califf, Richard Platt) concurred and added that, in their experience, design of a rigorous observational study was considerably more challenging than design of an RCT. They counseled that the required time, effort, and reiteration were often underestimated in the planning and design phases of observational studies.

The phenomenon of "P-value hunting" for data obtained from observational studies (and also from randomized studies) has been recently discussed by the American Statistical Association and it has led to the failure to reproduce results that were initially reported as "significant." The prespecification of models for observational studies may be a method to overcome this.

EFFECTIVENESS RESEARCH AND MODERN OBSERVATIONAL TRIALS

Trial designs that attempt to meld the robust reliability of traditional RCTs with the flexibility and generalizability of observational trials are already being used in surgical clinical research, and may guide the design of future neurosurgical studies.

Pragmatic Randomized Trials

Pragmatic randomized trials achieve high internal validity through the use of randomization, but are designed to be representative of general practice, and hence are readily generalizable. Pragmatic RCTs typically have a short list of patient eligibility criteria with relatively few exclusions, yielding a heterogeneous population representative of clinical practice. The trials are generally designed to integrate relatively smoothly into standard clinical care, so there are relatively few stipulations regarding concurrent medications or therapies. The outcome measures tend to be relatively few, simple, and objective (eg, mortality, disability) rather than complex scales, indices, or laboratory measures. Simple outcome measures facilitate performance in a wide range of practice settings. The design characteristics lead to a requirement for large sample sizes, so these trials usually include at least 1000 patients. Descriptors like large and simple are often used for pragmatic randomized trials. In principle, the pragmatic randomized trial can provide highly rigorous outcome data that can be readily generalized to the real-world needs of practicing physicians and their patients. A successful example in neurosurgery was the Corticosteroid Randomisation After Significant Head Injury (CRASH) trial, which recruited over 10 000 patients and demonstrated that corticosteroids were harmful in the treatment of severe head injury.[21]

Combined or Concurrent RCT and Observational Trials

Combined or Concurrent RCT and Observational Trials also seek high internal and external validity. If the RCT and observational study are concurrent but otherwise independent, clearly the first provides the proof of efficacy, while the second provides the generalizability. In other circumstances, the 2 are linked, and meant to be somewhat more mutually supportive. One approach is to enroll patients who are not eligible or agreeable to randomization in the observational study. This approach was used in the Spine Patient Outcomes Research Trial (SPORT) comparing lumbar discectomy to conservative care, although high rates of noncompliance with treatment assignment significantly decreased the power of the per protocol analysis.[22] In other cases, the separation of the RCT and observational trial might be based on physician willingness to participate in an RCT, or characteristics of the hospital or practice setting. These approaches are attractive in terms of productivity, but there are potential risks that the basis for assigning patients to the RCT or observational trial could generate biases which might alter outcomes of one or both studies.

Effectiveness in a Prospective Observational Trial

The Approaches and Decisions in Acute Pediatric TBI Trial (ADAPT; www.adapttrial.org) takes a different approach. Its chief mission is to collect data regarding the clinical care administered to children with severe head injury, the frequency with which different procedures and interventions are used, and the different ways in which they are combined. ADAPT has broad inclusion criteria and few exclusions, does not direct or specify interventions, and will capture data on 1000 patients from a wide array of medical centers. The study will rigorously examine the effectiveness of 6 commonly used interventions, such as continuous cerebrospinal fluid diversion, hyperosmolar therapy, or high caloric feeding. Formal primary and secondary hypotheses with well-defined, quantifiable outcome measures have been defined for each topic. For each hypothesis, analyses will be conducted to test the hypotheses. Confounded effects will be controlled for in the analyses using propensity scores derived from the observational data. Sample size estimates for the hypotheses informed the number of participants to be enrolled in the study. The approach is rigorous for assessment of effectiveness, while preserving the advantages of simplicity, flexibility, and generalizability from its observational design.

Embedding an RCT into a Prospective Observational Database

Large electronic health databases can be exploited as data collection tools that allow performance of RCTs with little additional effort. For example, the Thrombus Aspiration during ST-Segment Elevation Myocardial Infarction (TASTE) study comparing the outcomes of percutaneous coronary intervention with or without manual thrombus aspiration used the internet-based Swedish Coronary Angiography and Angioplasty Registry (SCAAR).[23] Randomization and all follow-up data collection were performed using the registry infrastructure and yielded a rigorous comparison of effectiveness in more than 7000 patients while imposing minimal additional workload on participating medical staff. The cost was about 10% that of a traditional RCT.

Performing a Virtual RCT in a Retrospective Observational Database

This is another way of describing an observational study conducted using the RCM in an existing large database. As an example, effectiveness of total mastectomy vs breast-sparing surgery was compared using RCM methods and the National Cancer Institute (NCI) Surveillance, Epidemiology, End Results (SEER) database. Even though the database contained 5-year survival data for only 5326 women, the study was able to demonstrate comparable survival between groups.[15] To identify the 30 variables that contributed to the propensity score estimate, the investigators conducted multiple interviews with surgeons and patients. The study successfully accomplished its goal, which was to determine whether published reports of comparable effectiveness from tertiary care medical centers would generalize to

community practice, while adding no extra burden to health care providers.

OBSERVATIONAL DATABASES AND REGISTRIES

Early progress in medical information technology was closely linked to the financial and administrative aspects of medical care, such as diagnostic coding, billing, and payments. Nonetheless, these same data can support insights into patient safety and outcomes and have been aggregated into large national databases to support clinical research. As EHRs and other digital data are integrated more fully into medical practice, there will be an explosion in the amount of information available on the experiences of thousands of neurosurgeons and millions of patients. The immediate challenge is to determine how this massive amount of data can best be captured and analyzed to provide reliable guidance for neurosurgical practice.

Mini-Sentinel and PCORnet: The National Patient-Centered Clinical Research Network are representative of current big data resources.[13,24,25] Mini-Sentinel was established in response to a congressional mandate for modernization of the FDA drug safety surveillance system. As of February 2015, the database included 48 million people and 358 million years of patient observation, contributed from partner organizations, which include insurance companies such as Aetna and Anthem, hospital chains such as HCA, and care organizations such as Kaiser Permanente. PCORnet is an initiative of the Patient-Centered Outcomes Research Institute (PCORI) that combines 11 clinical data research networks and 17 patient-centered data research networks, and supports the NIH Collaboratory Initiative. The majority of data comes from coding and billing databases, with a minor contribution from EHR data sources. To facilitate clinical research, a propensity score tool is available for Mini-Sentinel. Due to the data sources, the granularity and specificity of the data related to surgical procedures are limited. Therefore, the types of clinical research that could be performed are limited to practice patterns, utilization, and broad assessments of safety or outcomes for a class of procedures. Unfortunately, these databases lack much of the information required for many research topics in neurosurgery. Some patient characteristics may be inferred, but there are no data regarding physical findings, results of imaging or other laboratory studies, or severity of the diagnosed disorder. Data are not included on any specifics of the surgery or events that occurred during the operation. Much more detailed neurosurgical data would be required to guide neurosurgical practice.

In the short term, such data appear most likely to come from neurosurgery-specific registries and databases. Examples of registries relevant to neurosurgery include those supporting specialty certification, safety of FDA-approved devices, and longitudinal natural history studies. Although these serve valuable functions, they do not necessarily serve as good models for effectiveness research registries. Some of these registries are based on spontaneous reporting, whereas an effectiveness outcome registry

must be more complete and include all representative experience. Since a clinical research registry needs to contain all of the data that may prove necessary for eventual analyses, the number of data fields may be larger. The impact of missing and incomplete data is higher in studies testing specific hypotheses than in purely descriptive collections. Both the quality and accuracy of these data must be of the highest caliber to guide neurosurgical practice.

The creation of increasingly detailed national clinical data collection systems is being largely driven by the requirement for robust quality measurement, which has taken on a central role in the rapidly evolving health care landscape. In particular, clinical registries have seen explosive growth in recent years as methods to advance public and private patient safety initiatives and quality reporting mandates. Perhaps the most conspicuous recent promotion of registry use for public quality reporting was the authorization of Qualified Clinical Data Registries (QCDR) by Congress in 2014 (https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Qualified-Clinical-Data-Registry-Reporting.html). The QCDR is an alternative to traditional Physician Quality Reporting System (PQRS) methods that allow participants to satisfy PQRS requirements by reporting measures that have been developed and validated by the registry entity. Among all the available public reporting methods, QCDRs are particularly well suited to harness the power of registries in order to create disease- and treatment-specific measures that reflect realistic and relevant quality targets for neurosurgery and other medical specialties. In short, the national quality imperative is driving the creation of large and valuable information repositories that can be prospectively or retrospectively analyzed to further medical science.

Requirements for quality improvement have provided strong impetus for new registries, such as the NeuroPoint Alliance National Neurosurgery Quality and Outcomes Data Base (N^2QOD), a major initiative led by the American Association of Neurological Surgeons and supported by the American Board of Neurological Surgery, the Society of Neurological Surgeons, the Congress of Neurological Surgeons, and the Scoliosis Research Society.[26] The first project was the creation of an observational database for lumbar spine surgery, which, as of August 2015, included over 500 participating surgeons in 32 US states who had registered more than 20 000 patients. The database captures 52 "risk variables" and multiple outcomes, including the European Quality of Life Score (EQ-5D), Oswestry Disability Index (OSI), pain scales for limbs and back, and overall patient satisfaction. Participant compliance with data entry has been excellent, and initial data completeness and 12-month patient follow-up have exceeded 95% and 74%, respectively. Importantly, N^2QOD has demonstrated the ability of academic and community surgeons to work cohesively in producing the type of data-rich registry which could potentially support rigorous effectiveness research. One limitation of N^2QOD is that participants enter only selected cases (albeit through a standardized sampling methodology); rigorous research would require inclusion of all cases. "All care"

data collection is now being initiated in test N²QOD sites to determine the feasibility of this approach. Another limitation of N²QOD at the present time is the cost associated with the requirement for research coordinators to enter the data and verify accuracy. The N²QOD has also recently been approved as a QCDR, a designation which will raise the bar for complete data collection at participating sites.

Currently, N²QOD seeks to ensure data accuracy utilizing the following methods:

(1) Data cleaning: Routine reviews of data at the coordinating center ensure data completeness. Significant inconsistencies between data fields are also identified (eg, patients with very low disability and/or pain scores undergoing lumbar surgery; patients with simple pathologies such as disc herniation undergoing fusion, etc) and contributing centers are asked to reconcile unexpected variations.
(2) Self-audits: Contributing centers are periodically asked to review their data vs source documentation and report on the accuracy of reported information. Special emphasis is made on determining that submitted data reflect the eligible patient population, and that diagnostic and procedural categories are correctly recorded.
(3) Site audits: Contributing centers are occasionally subject to random site audits to ensure the accuracy of submitted data, particularly with respect to patient outcomes.

The following additional processes are planned to further improve data accuracy in N²QOD:

(1) Participating centers will be "rated" with respect to the quality of submitted data and their use of approved methods to ensure data accuracy.
(2) Third-party site audits will be initiated. These will include random and 'for-cause' audits.

In the long run, it appears inevitable that as EHR use becomes widespread, more detailed data on neurosurgery and other specialties will appear in the metadatabases. Eventually there may be a massive database available as a public utility to support clinical research and advance care in all aspects of medicine and surgery. In order to assure that neurosurgical data are appropriately represented in metadatabases, it is important that neurosurgeons participate and contribute to information technology forums and committees. Since much of this work occurs at the local level of hospital or academic institutions, there are many opportunities to become involved.

In addition to registries, clinical trials provide another source of data enriched with information important to neurosurgery, which could be used for RCM analyses or other research. Requirements for public sharing of data from federally funded clinical research will produce increasing numbers of such reservoirs. For example, NINDS will provide access to data from such studies as the Carotid Revascularization and Medical Management for Asymptomatic Carotid Stenosis (CREST-2)

trial (NCT02240862), ADAPT for pediatric traumatic brain injury, and Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) for adult traumatic brain injury (NCT02119182). To an increasing extent, industry is also making subject-level trial data available to qualified researchers, though sometimes these data are limited to participants in control groups.

OVERCOMING BARRIERS TO CLINICAL TRIALS

Equipoise

Achieving equipoise has been a challenge to the conduct of all RCTs, but is a particularly acute problem for surgical trialists. In extreme cases, a lack of equipoise on the part of patients or physicians has led to early termination of the clinical trial due to the inability to enroll patients in a reasonable time frame (eg, Radiosurgery or Open Surgery for Epilepsy trial (ROSE),[27] Early Randomized Surgical Epilepsy Trial (ERSET)[28]). Insufficient equipoise may lead the investigator to exclude patients judged to be manifestly good or poor candidates for the procedure under study, jeopardizing the demonstration of benefit and decreasing the ability to generalize outcomes. In practice, lack of equipoise is often regarded as a potential obstacle only for RCTs, and a common method of addressing the obstacle is to propose an observational study. In reality, unless there is some variation in practice within the population being studied, a proper observational study cannot be conducted either. No statistical method can generate a meaningful comparison between cases that are truly "apples and oranges."

Equipoise is typically defined as a genuine uncertainty among clinical experts as to which of 2 treatment choices will be best for the patient. The functional meaning and significance of equipoise depend to some extent on the locus of perception: as one author asked, "whose equipoise is it anyway?."[29] In a classic article, Freedman drew attention to 2 concepts of equipoise that have shaped much subsequent thinking.[30] In one sense, equipoise might be considered a lack of consensus judgment within the neurosurgical community, such as between clipping and coiling for intracranial aneurysms, or anterior vs posterior decompression for cervical spondylotic myelopathy. Community equipoise is reflected by the observed measure of "variation in practice." In a second sense, equipoise may be perceived as a specific clinician's perception of 2 treatment choices for a given patient (Freedman's "clinical equipoise"). Freedman's point was that no physician, faced with a specific, real patient, can be poised on an academician's knife-edge of uncertainty, and that recognition of community variation in practice was a more important fact justifying an RCT from an ethical standpoint. Clinician lack of equipoise can be described positively as "expert judgment," or less positively as "physician bias." Finally, equipoise at the level of the patient is a third concept not directly addressed by Freedman. Lack of equipoise at this level is termed "patient preference." All of these are expressions of equipoise, though they are clearly

different, distinguishable concepts that have different implications for clinical trial design and conduct.

Pragmatically speaking, any experimental clinical trial comparing treatments requires the existence of sufficient community equipoise to motivate investigators to participate and other physicians to refer prospective patients into the trial centers. An RCT can only be carried out if clinician and patient equipoise exist for at least as many patient/physician pairs as the trial design requires to complete accrual. Typically, clinical and patient equipoise is left entirely to the judgment of the individuals. A recent innovation is the "equipoise panel," a means to obtain and apply a broader sample of expert opinion. In the SLIP study of spinal fusion, a panel of 10 expert surgeons reviewed each case and provided their opinion as to whether or not equipoise existed for randomization of that individual. After institution of the Equipoise Panel, the proportion of patients agreeing to randomization increased from 40% to 81%.[31] Other types of educational or informational interventions might also be appropriate in assisting investigators and patients in their determinations regarding equipoise.

Alternative types of randomization procedures may allow equipoise to be achieved more readily. In expertise-based randomized trials, the expert surgeon is recognized as a key component of the treatment under study, and study surgeons would typically only perform one of the 2 treatments under study.[32] Participants are randomized to surgeon, and hence treatment, as a unit. Since the treatment is carried out by a surgeon who is expert in that procedure and believes in its efficacy, equipoise is bypassed for trial physicians, though challenges may remain for the patient. In cluster-based randomization, clusters or groups of participants, such as patients in a particular medical practice or a particular city, are randomly assigned to treatments, that is, all individuals in the cluster receive the same treatment. For both physicians and patients, equipoise is reduced to acceptance of the proffered treatment. Although these alternative approaches to randomization may increase the feasibility of a clinical trial, they may inadvertently introduce bias. Cluster analysis has the drawback that the analysis may be more complex, and within-cluster correlations can lead to reduced statistical power.

Observational studies may exploit variation in routine clinical practice to produce a treatment comparison. The comparison may be based on variation in the treatments assigned by individual physicians, or among groups, such as hospitals or geographic regions. Individual clinician equipoise is not an issue in observational studies, but community equipoise is. Without naturally occurring variation in practice, all patients in the relevant population will receive the same treatment, and no treatment comparison can be done. Although superficially the treatment comparison may resemble that from an RCT using expertise or cluster-based randomization, the assignment is not randomized, and therefore the results from such observational data are vulnerable to selection bias. This bias can be controlled using an approach such as propensity score analysis, providing the drivers for clinical decision making are well understood and have been captured adequately in the observational database.

Community equipoise is also relevant to trial planning to the extent that the trial outcome will be accepted by the neurosurgical community and lead to a consequent change in clinical practice. For a truly novel therapy where there is no prior opinion, one definitive trial can have a major and rapid impact on practice. An example is the phase III trial of temozolomide for glioblastoma, which was followed almost immediately by regulatory approval and rapidly ushered in a new era in management.[33] On the other hand, when strong opinion exists within a community on the part of either patients or physicians, even clinical trials with clear answers may fail to influence practice. Examples include the continued performance of vertebroplasty by some interventional radiologists after 2 RCTs had demonstrated that it was not efficacious[34] or the failure of a large panel of RCTs to convince certain elements of American society of the safety of mumps, measles, and rubella vaccination.[35] From this perspective, community equipoise is critical to all clinical trials, whether formal RCTs, descriptive registry trials, or modern observational trials.

The Human Factor—The Neurosurgical Clinical Research Workforce

Neurosurgeons are under great pressure to perform patient care, whether in the clinic, hospital, or operating room. There is widespread recognition that clinical research is important for continued improvement in evidence-based neurosurgery; however, this does not necessarily translate into protected time or resources for clinical research.[36] Practical support from administrators in academic and large private institutions by provision of time, salary support, or other resources, such as access to clinical space, would significantly reduce the barriers to performance of neurosurgical research. Clinical science is a team effort, and it should be recognized that protection of personnel means not only neurosurgeons but also research nurses, information technology experts, and other contributors. Provision of more education and training about clinical research to residents, fellows, junior faculty, and practicing neurosurgeons would significantly increase the pool of individuals who realize that research can be exciting, intriguing, enjoyable, and fulfilling, and who therefore would participate as investigators in research.[37]

Rapid innovations in technology could greatly change the experience of clinical research and increase participation by the neurosurgery community. User friendly data portals decrease the time burdens of data entry and make participation accessible to more clinicians. Several workshop participants shared their vision of a future in which clinical research was fundamentally integrated into the daily routine of neurosurgical practice, and evidence-based personalized medicine flourished naturally with minimal effort. Some specialties have advanced somewhat more rapidly than neurosurgery into this future. For example, registry participation approaches 98% in thoracic surgery[38] and for more than 20 years over 70% of pediatric oncology patients have

been entering prospective clinical cancer trials.[39] Neurosurgeons could perhaps learn from the experiences of these subspecialties to advance to a more fully participatory "Science of Practice."

Finally, in all of this, one must not forget the patient. In fact, involving patients at all stages of study design is critical for success of an RCT: patients can guide determinations of question relevance, protocol feasibility, and outcomes of interest. A specific example of this patient involvement is the recent appreciation of the patient-reported outcomes, where patients themselves provide answers to and report on variables related to outcomes. As this type of analysis expands, it would be important for patients and patient advocates to be involved in the design of studies in order to incorporate what is meaningful to them.

CONCLUSIONS AND RECOMMENDATIONS

RCTs have the highest internal validity, provide the most rigorous demonstration of efficacy, and are established as the most reliable level of evidence for guiding practice. In order for RCTs to be performed more frequently in neurosurgery, one of the barriers which must be overcome is achievement of equipoise. Discussion at this workshop provided clarity and recognition that there is not a "single" equipoise but rather multiple constructs. Several technical approaches such as equipoise panels, expertise-based randomization, and cluster randomization provide tools for dealing with the challenges of equipoise. A future workshop on equipoise might be considered to guide neurosurgical opinion and examine other means of achieving equipoise in randomized studies.

Many neurosurgeons are dissatisfied with the limited external validity of traditional RCTs; they want data relevant to heterogeneous patient populations and practice settings. Effectiveness research provides theoretical constructs and concrete examples of methods for performance of clinical research which addresses "real-world" needs. Registries and other observational databases can serve as a vehicle for effectiveness research, with potential advantages being cost efficiency and minimal burden to participating physicians.

Advances in the application of information technology to surgical practice can lead to a future in which clinical research is firmly embedded into the daily practice of neurosurgery. Data collection will be widespread, detailed, and inclusive, creating extraordinarily rich, ever-evolving databases. Analysis will be ongoing, feeding back relevant recommendations and guidance for best practices to all neurosurgeons, in a manner specific to their particular setting and patients. The routine collection and analysis of patient characteristics, processes of care, and outcomes—inseparable from clinical practice—will guide the evolution and optimization of neurosurgical practice. Information technology can also improve the conduct of RCT.

Traditionally, a very small percentage of physicians have produced the vast majority of new medical knowledge, while the others have served as knowledge consumers. This reliance on a small scientific elite for the generation of novel health care information is giving way to a requirement that all physicians engage in scientific inquiry through the acquisition and analysis of practice data. In the near future, the majority of physicians will work together to collect and analyze clinical data to define specialty-wide standards for health care value, safety, and effectiveness. This will necessitate a new methodology for clinical research based in real-world clinical practice called the "science of practice."

Three key features define the science of practice: (1) habitual and systematic collection of data, inseparable from clinical activity; (2) analysis of practice data to generate new knowledge; and (3) the application of that knowledge to produce iterative advancement in health care. The science of practice algorithm is a registry-based approach to clinical research. If the registries are designed carefully, the data will support causal inferences about the efficacy and effectiveness of surgical interventions. Steps taken today provide the groundwork for the attainment of this vision.

Recommendations were made for some specific tools that are needed for information technology applications:

- Widely accepted and utilized common data elements for neurosurgical indications and procedures.
- A unique national patient identifier, which could be used to track an individual across multiple registries and studies while preserving privacy, such as the Global Unique Identifier (GUID; https://fitbir.nih.gov/jsp/contribute/guid-overview.jsp).
- Consensus on critical data fields that should be captured and extracted from EHR.
- Methods for inclusion of actual digital imaging data rather than reliance on text extracts from radiology reports.
- Covariates that determine treatment assignment need to be defined and measured in order to draw causal inferences about treatment.
- Inclusion of patients and patient advocates in study design, in order to ensure that outcomes that are meaningful to them are measured.

Neurosurgeons might engage with other communities, such as thoracic surgery, where registry use is more widespread and established, for lessons learned.

The hurdles for data detail, quality, and completeness are extremely high for observational databases meant to successfully support reliable, rigorous effectiveness research. Experience will be required to guide design and implementation of information technologies to support the needs of neurosurgery research. Methods need to be developed for ongoing auditing of submitted data for completeness and accuracy.

Important initiatives that incorporate these principles, such as N^2QOD, are already underway.

Timely performance of clinical trials on selected topics would provide another opportunity for experience supporting

2 components of the envisioned "Science of Practice": (1) design and implementation of a registry which supports a rigorous effectiveness study; and (2) acceptance and incorporation of the results in neurosurgical practice. Therefore, a recommendation was made for a follow-up workshop to focus on specific topics and designs for neurosurgical clinical trials which would include a registry and apply contemporary approaches to observational and effectiveness research.

A second potential meeting might bring together stakeholders such as neurosurgeons, professional associations, advocacy groups, insurers, and others to create specific recommendations and means for advancing the vision of a science of practice. Such a meeting would specifically address information technology and creation of a general, inclusive neurosurgery registry.

Disclosures

REFERENCES

1. Kiehna EN, Starke RM, Pouratian N, Dumont AS. Standards for reporting randomized controlled trials in neurosurgery. *J Neurosurg*. 2011;114(2):280-285.
2. Mansouri A, Cooper B, Shin SM, Kondziolka D. Randomized controlled trials and neurosurgery: the ideal fit or should alternative methodologies be considered? *J Neurosurg*. 2016;124(2):558-568.
3. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials*. 2009;10:37.
4. Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med*. 1986;15(5):451-474.
5. Yabroff KR, Harlan L, Zeruto C, Abrams J, Mann B. Patterns of care and survival for patients with glioblastoma multiforme diagnosed during 2006. *Neuro Oncol*. 2012;14(3):351-359.
6. Ahn JM, Kang SJ, Yoon SH, et al. Meta-analysis of outcomes after intravascular ultrasound-guided versus angiography-guided drug-eluting stent implantation in 26,503 patients enrolled in three randomized trials and 14 observational studies. *Am J Cardiol*. 2014;113(8):1338-1347.
7. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286(7):821-830.
8. Li Q, Zhang Z, Yin RX. Drug-eluting stents or coronary artery bypass grafting for unprotected left main coronary artery disease: a meta-analysis of four randomized trials and seventeen observational studies. *Trials*. 2013;14:133.
9. Florell RC, Macdonald DR, Irish WD, et al. Selection bias, survival, and brachytherapy for glioma. *J Neurosurg*. 1992;76(2):179-183.
10. Gutin PH, Prados MD, Phillips TL, et al. External irradiation followed by an interstitial high activity iodine-125 implant "boost" in the initial treatment of malignant gliomas: NCOG study 6G-82-2. *Int J Radiat Oncol Biol Phys*. 1991;21(3):601-606.
11. Selker RG, Shapiro WR, Burger P, et al. The Brain Tumor Cooperative Group NIH Trial 87-01: a randomized comparison of surgery, external radiotherapy, and carmustine versus surgery, interstitial radiotherapy boost, external radiation therapy, and carmustine. *Neurosurgery*. 2002;51(2):343-355; discussion 355-347.
12. Neukamp M, Perler G, Pigott T, Munting E, Aebi M, Roder C. Spine Tango annual report 2012. *Eur Spine J*. 2013;22(suppl 5):767-786.
13. Psaty BM, Breckenridge AM. Mini-Sentinel and regulatory science–big data rendered fit and functional. *N Engl J Med*. 2014;370(23):2165-2167.
14. Rubin D. For objective causal inference, design trumps analysis. *Ann Appl Stat*. 2008;2:808-840.
15. Rubin D. Statistical inference for causal effects, with emphasis on applications in epidemiology and medical statistics. *Handbook of Statistics* 2008;27:28-63.
16. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 2016; in press.
17. Torner JC, Zhang J, Piepgras D, et al. Comparative effectiveness of long-term outcomes of treatment of unruptured intracranial aneurysms. *Stroke*. 2012; 42:A39.
18. Cawley J. A selective review of the first 20 years of instrumental variables models in health-services research and medicine. *J Med Econ*. 2015;18(9): 721-734.
19. Yang S, Eaton CB, Lu J, Lapane KL. Application of marginal structural models in pharmacoepidemiologic studies: a systematic review. *Pharmacoepidemiol Drug Saf*. 2014;23(6):560-571.
20. Hogan JW, Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Stat Methods Med Res*. 2004;13(1):17-48.
21. Roberts I, Yates D, Sandercock P, et al. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. *Lancet*. 2004;364(9442):1321-1328.
22. Lurie JD, Tosteson TD, Tosteson A, et al. Long-term outcomes of lumbar spinal stenosis: eight-year results of the Spine Patient Outcomes Research Trial (SPORT). *Spine (Phila Pa 1976)*. 2015;40(2):63-76.
23. Frobert O, Lagerqvist B, Gudnason T, et al. Thrombus Aspiration in ST-Elevation myocardial infarction in Scandinavia (TASTE trial). A multicenter, prospective, randomized, controlled clinical registry trial based on the Swedish angiography and angioplasty registry (SCAAR) platform. Study design and rationale. *Am Heart J*. 2010;160(6):1042-1048.
24. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood)*. 2014;33(7):1178-1186.
25. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc*. 2014;21(4):578-582.
26. Asher AL, Speroff T, Dittus RS, et al. The National Neurosurgery Quality and Outcomes Database (N2QOD): a collaborative North American outcomes registry to advance value-based spine care. *Spine (Phila Pa 1976)*. 2014;39(22 suppl 1):S106-S116.
27. Barbaro NM, Quigg M, Broshek DK, et al. A multicenter, prospective pilot study of gamma knife radiosurgery for mesial temporal lobe epilepsy: seizure response, adverse events, and verbal memory. *Ann Neurol*. 2009;65(2):167-175.
28. Engel J, Jr, McDermott MP, Wiebe S, et al. Early surgical therapy for drug-resistant temporal lobe epilepsy: a randomized trial. *JAMA*. 2012;307(9):922-930.
29. Lilford RJ. Ethics of clinical trials from a bayesian and decision analytic perspective: whose equipoise is it anyway? *BMJ*. 2003;326(7396):980-981.
30. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*. 1987;317(3):141-145.
31. Ghogawala Z, Schwartz JS, Benzel EC, et al. Increased patient enrollment to a randomized surgical trial through equipoise polling of an expert surgeon panel. *Ann Surg*. 2015.
32. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ*. 2005;330(7482):88.
33. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987-996.
34. Buchbinder R, Osborne RH, Kallmes D. Invited editorial presents an accurate summary of the results of two randomised placebo-controlled trials of vertebroplasty. *Med J Aust*. 2010;192(6):338-341.
35. Taylor LE, Swerdfeger AL, Eslick GD. Vaccines are not associated with autism: an evidence-based meta-analysis of case-control and cohort studies. *Vaccine*. 2014;32(29):3623-3629.
36. Suliburk JW, Kao LS, Kozar RA, Mercer DW. Training future surgical scientists: realities and recommendations. *Ann Surg*. 2008;247(5):741-749.
37. Ko CY, Whang EE, Longmire WP, Jr, McFadden DW. Improving the Surgeon's participation in research: is It a problem of training or priority? *J Surg Res*. 2000;91(1):5-8.
38. Shahian DM, Jacobs JP, Edwards FH, et al. The society of thoracic surgeons national database. *Heart*. 2013;99(12):1494-1501.
39. Tejeda HA, Green SB, Trimble EL, et al. Representation of African-Americans, Hispanics, and whites in National Cancer Institute cancer treatment trials. *J Natl Cancer Inst*. 1996;88(12):812-816.

Acknowledgments

COMMENT

This manuscript presents a detailed summary of the findings of a workshop sponsored by the National Institute of Neurological Disorders and Stroke (NINDS). The manuscript reflects the current status of clinical investigation in the field of neurosurgery. The focus of the findings by the authors is on the area of effectiveness research. Rather than presenting the results of prior clinical investigations, the authors propose goals and guidelines for future studies. The recommendations are to direct future research in the field of neurosurgery.

The authors correctly point out the enormous difficulty involved in successfully planning and executing randomized clinical trials in neurosurgery. The limitations of randomized clinical trials are well known to every neurosurgeon. Upon completion, we are not infrequently left with more questions than answers regarding "best approaches" to clinical care. The authors therefore offer methods in which observational studies might be more realistically and successfully completed in order to advance the field of neurosurgery. Most importantly, the authors propose the requirements for the rigorous demonstration of effectiveness from observational data. These guidelines are essential information for clinical investigations who wish to perform effectiveness research using current observational trials.

The barriers to randomized clinical trials in our field are greater now than ever before. Given the challenges of achieving clinical equipoise in the vast majority of neurosurgical trials, the authors conclude that clinical registries are a far more pragmatic approach to answering the questions necessary to appropriately guide practice in neurosurgery. The time is past where a very small percentage of surgeons produce the majority of new medical knowledge. As the authors correctly conclude, "In the near future, the majority of physicians will work together to collect and analyze clinical data to define specialty-wide standards for health care value, safety and effectiveness."

Peter Gerszten
Pittsburgh, Pennsylvania