Teaching Evidence Based Medicine to Neurosurgery Residents

Robert E. Harbaugh, MD, FAANS, FACS, FAHA
Director, Institute of the Neurosciences
University Distinguished Professor & Chair, Department of Neurosurgery
Professor, Engineering Science and Mechanics
Penn State University

Disclosures

Active Grant Funding
Integra Foundation, NIH - R01-NS049135-01 and R01-HL083475-01A2

Consultant
Micromechatronics, Inc., Advance Medical, Inc.

Stock
Micromechatronics, Inc., Advance Medical, Inc., Cortex, Inc.

Fiduciary Responsibility
President, CHYNA, LLC
Chief Medical Officer, NuHope, LLC
Treasurer, NeuroPoint Alliance, Inc.

U.S. Patent Applications
20060212097 and 20070138915

Two Themes and One Question

Theme 1. Interpreting the neurosurgery literature and performing good clinical research requires an understanding of clinical trial design and biostatistics.

Theme 2. These skills are *vitally important* for practicing neurosurgery in the 21st Century.

Question 1. How can we best teach our residents these skills?

Evidence Based Medicine

EBM is an algorithm for clinical decision making that…

Uses clinical research published in peer reviewed journals as the primary or only source of evidence.

Defines a hierarchy of strength of evidence based on the methodology of data collection and analysis - with RCTs accepted as the gold standard for clinical research.

Derives standards, guidelines and options from a methodologically rigorous analysis of evidence using this hierarchy.

Determines the quality of healthcare by adherence to established practice parameters.

Evidence Based Medicine

Rationale:

Application of the EBM algorithm will improve patient care by fostering better clinical decision making.

Decisions will be made on the basis of scientifically valid data rather than intuition, training or other non-verifiable factors.

Better clinical decision making will result in improved outcomes.

Understanding the EBM Algorithm for Neurosurgery

Why are RCTs considered a gold standard?

All clinical studies are prone to error from bias, confounding and chance

The RCT is accepted as the gold standard because, ideally, this design minimizes errors from bias, confounding and chance

- Prospective - decreases information and recall bias
- Randomized and controlled - equally distributes confounders and eliminates internal selection bias
- Statistical analysis (P value and Power calculations) limits type I and type II chance errors

The Ideal RCT

The ideal RCT has three essential components:

1. Concurrent comparison to eliminate temporal bias

2. Objective observation of clear and clinically meaningful end-points to eliminate observer bias and to assure correlation of statistical and clinical significance

3. Randomization of a representative population of adequate size to equally distribute confounders and reduce the chance for sampling errors

A double-blind study of aspirin following myocardial infarction, with death as the outcome of interest, would approach the ideal

Problems with Surgical RCTs

Ideal RCT - Essential component #1 - Concurrent comparison

Surgical RCTs - Intention to treat analysis and crossovers

To preserve the benefits of randomization it is necessary to analyze patients in their assigned groups even if they crossover to another treatment arm.

Is this rational if "surgical" patients never receive surgery and "non-surgical" patients do? This is particularly problematic because the crossover periods are often asymmetrical.

Statistical methods exist to ameliorate this problem but they do not eliminate it

Problems with Surgical RCTs

Spine Patient Outcomes Research Trial (SPORT)
JAMA 296:2441-2450, 2006 - Lumbar Disc Herniation

Of 256 patients assigned to nonoperative care, 30 % had surgery at 3 months and 45% had surgery within 2 years.

The *as-treated analyses* demonstrated "strong, statistically significant advantages… for surgery at all follow-up times," the intent to treat analysis revealed only slightly better outcome scores for the group assigned to surgery at each time point.

Based on the *intention to treat analysis*, outcomes with surgery were not statistically significantly better than outcomes with non-surgical treatment.

What was reported?

"Study Questions Need to Operate on Disk Injuries"

New York Times, November 22, 2006

"Everyone was hoping the study would show which was better ...
...and everyone was surprised by the tremendous number of crossovers... In the end, the study could not provide definitive results on the best course of treatment because so many patients chose not to have the treatment that they had been randomly assigned."

Problems with Surgical RCTs

Ideal RCT - Essential component #2 - Objective observation of clear endpoints

Surgical RCTs - Blinding is difficult or impossible and endpoints may be vague, introducing bias for or against surgery

Methodological Concerns: Objective Observation of Clear Endpoints

International Subarachnoid Aneurysm Trial (ISAT)
Lancet 360: 1267-1274, 2002.

Functional health status reporting

Are the differences between adjacent categories on the Modified Rankin Scale clinically significant?

Can a postal questionnaire differentiate between “no symptoms” & “a few symptoms” or between “moderate” & “moderately severe” disability?

Is the method for determining functional health status biased in favor of coiling?

ISAT Functional Health Status at One Year

MRS Score	Questionnaire Response	Coiling (N=801)	Clipping (N=793)	P value
0	I have no symptoms and I cope well with life.	207	152	<0.05
1	I have a few symptoms but these do not interfere with my everyday life.	217	220	NS
2	I have symptoms which have changed my life but I am still able to look after myself.	187	178	NS
3	I have symptoms which have significantly changed my life and prevent me from coping fully, and I need some help looking after myself.	80	106	NS
4	I have quite severe symptoms which mean I need to have help from other people but I am not so bad as to need attention day and night.	24	32	NS
5	I have major symptoms which severely handicap me and I need constant attention day and night.	21	25	NS
6		65	80	NS
****	***************************************	*******	*******	*****
3-6		190	243	<0.05
4-6		110	137	NS
2-6		377	421	NS

Methodological Concerns:
Objective Observation of Clear Endpoints

Is the method of obtaining functional health status outcomes biased in favor of coiling?

Yuval et al. Arch of Int Medicine 160: 1142-1146, 2000

In unblinded studies patients report better outcomes for treatments that are newer

Ford/Mercedes comparisons

Problems with Surgical RCTs

Ideal RCT - Eessential component #3 - Randomization of an adequately sized, representative population

Surgical RCTs - The surgeon has an implicit contract with the patient to offer the best care available (*therapeutic imperative)*.

If surgeons do not believe that both treatment arms are equally efficacious (*equipoise*) they will offer surgery to patients most likely to benefit and randomize those least likely to benefit.

Randomizing patients unlikely to benefit from treatment is analogous to performing a diagnostic test on patients unlikely to have a disease.

Populations & Interpretation of Studies

Populations

Target - Population to which the study should apply

Accessible - Subset of target population available for study

Intended - Subset of accessible population who are sampled

Actual - Subset of intended population enrolled in the study

How closely the actual population mimics the target population will determine the generalizability of the study

Diagnostic Tests and the Likelihood of Disease

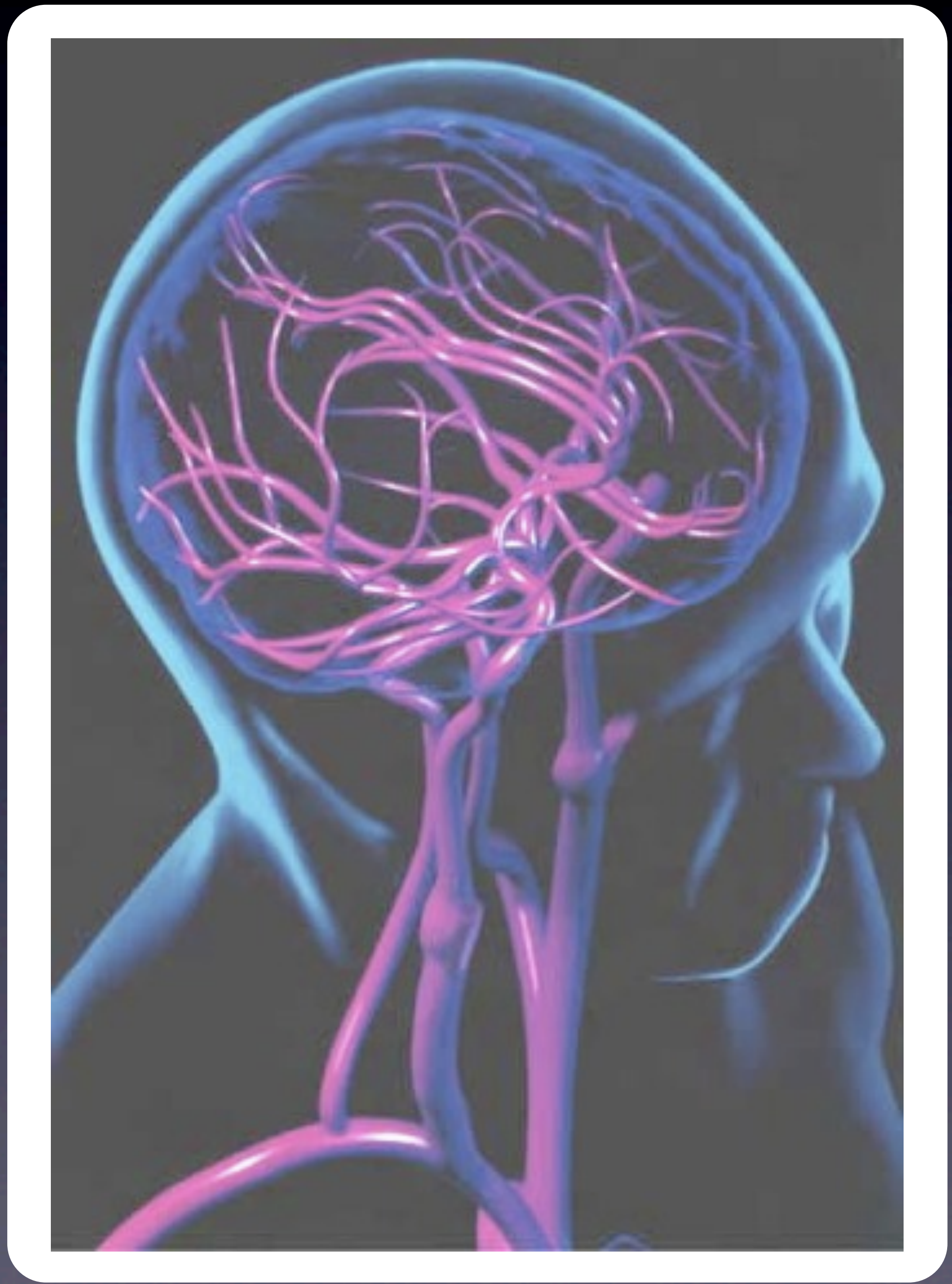
Baye's Theorem

Thomas Bayes (1702-1761) was an English theologian and mathematician who investigated the probability of an event, given a condition to the probability of a condition, given an event.

Applied to diagnostic tests, Bayes' theorem demonstrates that *to know the predictive accuracy of a test result it is necessary to know not only the sensitivity and specificity of the test but also* <u>*the prior probability of disease in the population being tested*</u>

Prior Probability and Diagnostic Tests

- Assume a diagnostic test for carotid stenosis >60% that has a specificity of .95 and a sensitivity of .80
- The test is given to two groups of 1000 patients
  - High Risk - Elderly smokers with repetitive TIAs (pr = .90)
  - Low Risk - Young, asymptomatic athletes (pr = .01)



Prior Probability and Diagnostic Tests

Results in 1000 high risk patients (se-.80 / sp-.95 / pr-.9)

	Positive test	Negative test	Actual
Stenosis	**720**	**180**	**900**
No Stenosis	**5**	**95**	**100**
Total	**725**	**275**	**1000**

PPA = 720/725 or *.993* NPA = 95/275 or .345

The post test probability that a patient with a positive test actually has >60% stenosis is *99.3%*.

Prior Probability and Diagnostic Tests

Results in 1000 low risk patients (se-.80 / sp-.95 / pr-.01)

	Positive test	**Negative test**	**Actual**
Stenosis	**8**	**2**	**10**
No Stenosis	**50**	**940**	**990**
Total	**58**	**942**	**1000**

PPA = 8/58 or *.138* **NPA = 940/942 or *.998***

The post test probability that a patient with a positive test actually has >60% stenosis is *13.8%*.

Analogy:
Diagnostic Tests and Clinical Trials

Diagnostic tests

1. Absence of disease

   Negative test result

2. Presence of disease

   Positive test result

3. Sensitivity

4. Specificity

5. Prior probability of disease (patient selection) profoundly affects the predictive value of the test

Analogy:
Diagnostic Tests and Clinical Trials

Diagnostic tests

1. Absence of disease

   Negative test result

2. Presence of disease

   Positive test result

3. Sensitivity
4. Specificity
5. Prior probability of disease (patient selection) profoundly affects the predictive value of the test

Clinical trials

1. Absence of Rx effect

   Negative study result

2. Presence of Rx effect

   Positive study result

3. Power
4. 1- P value
5. Prior probability of Rx effect (patient selection) profoundly affects the predictive value of the study

CEA and EBM

The evaluation of CEA is often cited as a prime example of the application of the principles of EBM

CEA was done for 30 years without good evidence of efficacy

RCTs were done to determine the efficacy of CEA for stroke prevention in patients with symptomatic and asymptomatic carotid stenosis

The published results of these RCTs and guidelines derived from them had a significant effect on patient care

CEA and EBM

Summary of CEA RCTs

NASCET / ECST et al. - There is a marked benefit of CEA for stroke prevention in symptomatic patients with >70% stenosis, a moderate benefit in patients with 50-70% stenosis and no benefit in patients with less than 50% stenosis

ACAS / ECST et al. - There is a moderate benefit of CEA for stroke prevention in asymptomatic patients with >60% stenosis

Statistical Significance is not Clinical Significance

Stroke Prevention by CEA at Two Years

Symptomatic Patients (NASCET)

Stenosis	Absolute RR	Relative RR	NNT	P-Value
70-99%	17.0%	67%	6	0.000051
50-69%	6.5%	29%	15	0.045
30-49%	3.8%	20%	26	0.16

Asymptomatic Patients (ACAS)

Stenosis	Absolute RR	Relative RR	NNT	P-Value
>60%	1.5%	30%	67	<.05 (at 5 years)

Comparative Effectiveness Research
vs
Personalized Medicine

Comparative Effectiveness Research (CER)

The direct comparison of health care interventions to determine which have the greatest benefits, harms and costs

Personalized Medicine

A medical model proposing the customization of healthcare, with all decisions and practices being tailored to the individual patient by use of genetic and other information

Null Hypothesis Testing Studies

In NHT, clinical trials are performed to determine the presence or absence of a treatment effect - i.e. to reject or accept the null hypothesis - for the treatments being studied - perfect for CER

The null hypothesis states that there is no difference between treatments.

The null hypothesis must be either “true” or “false”.

The “probability that a treatment works” is a meaningless statement in NHT.

The magnitude of a treatment effect, if any, is assumed to be constant.

Are There Alternatives to NHT?

NHT is not the only option for clinical research.

Analysis of audited registries could be used more extensively.

Such analysis differs from NHT in several important ways.

Registries

An audited, prospective registry with adequate risk stratification, evaluation of processes of care and meaningful clinical outcomes is a very powerful tool that overcomes many of the limitations of RCTs.

Analysis of registry data would allow us to determine best practices and refine surgical indications.

Registries are well adapted to an iterative process of surgeon specific QI and personalized medicine.

Multi-specialty registries could allow the best features of CER and Personalized Medicine to be combined.

Two Themes and One Question

Two Themes and One Question

Theme 1. Interpreting the neurosurgery literature and performing good clinical research requires an understanding of clinical trial design and biostatistics.

Two Themes and One Question

Theme 1. Interpreting the neurosurgery literature and performing good clinical research requires an understanding of clinical trial design and biostatistics.

Theme 2. These skills are vitally important for practicing neurosurgery in the 21st Century.

Two Themes and One Question

Theme 1. Interpreting the neurosurgery literature and performing good clinical research requires an understanding of clinical trial design and biostatistics.

Theme 2. These skills are vitally important for practicing neurosurgery in the 21st Century.

Question 1. How can we best teach our residents these skills?

Thank You for Your Attention