Why is our literature contradictory?

Sachin Kheterpal, MD MBA
Associate Professor
Vice Chair, Clinical Affairs & Quality

Department of Anesthesiology
University of Michigan
Disclosures

• No financial, consulting, contractual relationships with any vendor

• I left clinical medicine for “Corporate America” (General Electric)
  – People called me a genius

• I came back to clinical medicine
  – I’m an idiot
Overview

- Perioperative outcomes
- Research gold standard
- Efficacy vs effectiveness
- Observational research
- Reproducibility
- Multicenter collaboration
Perioperative morbidity & mortality

• Anesthesia events
  – Mortality: 1 in 200,000
  – Difficult airway: 2-4%
  – Visual loss: 0.1 – 1%

• Organ system major morbidity
  – Acute kidney injury: 1 – 3%
  – Myocardial infarction: 1 – 3%
  – Stroke: 1%
  – Acute Lung Injury: ??

• We are focused on the needle in the haystack
Perioperative period

- Physically intrusive intervention
- Risky and expensive
- Very difficult to blind study individuals
- Challenging to ethically randomize
  - Difficult airway, Hypotension, Hypertension
- “Random” clinical decisions (RCDs) rampant
  - Wide variation in practice because of few guidelines
- Challenging to recruit specific populations of interest
  - Pediatrics
  - Emergency surgery w/ non-optimized patient
The Research Gold Standard

- Prospective Randomized Controlled Trial (RCT)
  - Placebo
  - Blinded

- Strongest evidence
- Detailed protocols
- “Eliminate” alternate causality
- Power analysis → prospective study size
Not so golden

Infrequent events → large study

$\text{\$\$\$ / patient} \rightarrow \text{small study}

- **Controlled** trial is not **routine** clinical practice
- Specific, small study extrapolated to population at large
Aprotinin – small N

- Randomized 22 redo sternotomy patients to aprotinin or control
- EBL: 286 ML vs 1509 ML
- “No adverse cardiorespiratory effects were seen”
Aprotinin – bigger N

Full-Dose Aprotinin Use in Coronary Artery Bypass Graft Surgery: An Analysis of Perioperative Pharmacotherapy and Patient Outcomes

- Meta-analysis of double blind RCTs in cardiac surgery
- 1,723 cases across primary and redo sternotomy
- All component trials sponsored by Bayer
- Outcomes
  - Mortality, cerebrovascular event, MI, vasopressor use
  - Did NOT evaluate renal failure
- Noted an IMPROVEMENT in all outcomes
The floodgates open
Like most veterinary students, Doreen breezes through Chapter 9.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broken leg</td>
<td>Shoot</td>
</tr>
<tr>
<td>Infected eye</td>
<td>Shoot</td>
</tr>
<tr>
<td>Splayed hoof</td>
<td>Shoot</td>
</tr>
<tr>
<td>Runny nose</td>
<td>Shoot</td>
</tr>
<tr>
<td>Fever</td>
<td>Shoot</td>
</tr>
<tr>
<td>Open sores</td>
<td>Shoot</td>
</tr>
<tr>
<td>Closed sores</td>
<td>Shoot</td>
</tr>
<tr>
<td>Swollen belly</td>
<td>Shoot</td>
</tr>
<tr>
<td>Ornery</td>
<td>Shoot</td>
</tr>
<tr>
<td>Swayback</td>
<td>Shoot</td>
</tr>
<tr>
<td>Erratic heart</td>
<td>Shoot</td>
</tr>
<tr>
<td>Hearing loss</td>
<td>Shoot</td>
</tr>
<tr>
<td>Bad breath</td>
<td>Shoot</td>
</tr>
<tr>
<td>Mane mange</td>
<td>Shoot</td>
</tr>
<tr>
<td>Toothache</td>
<td>Shoot</td>
</tr>
<tr>
<td>Nervousness</td>
<td>Shoot</td>
</tr>
<tr>
<td>Dentistry</td>
<td>Shoot</td>
</tr>
</tbody>
</table>
Aprotinin

- Case reports of renal dysfunction and stroke
- Controversy ensues…

A propensity score case-control comparison of aprotinin and tranexamic acid in high-transfusion-risk cardiac surgery

Keyvan Karkouti, W. Scott Beattie, Kathleen M. Datillo, Stuart A. McCluskey, Mohammed Ghannam, Ahmed Hamdy, Duminda N. Wijeysundera, Ludwik Fedorko, and Terrence M. Yau

The Risk Associated with Aprotinin in Cardiac Surgery

Dennis T. Mangano, Ph.D., M.D., Iulia C. Tudor, Ph.D., and Cynthia Dietzel, M.D., for the Multicenter Study of Perioperative Ischemia Research Group and the Ischemia Research and Education Foundation*
The final word?

A Comparison of Aprotinin and Lysine Analogues in High-Risk Cardiac Surgery

• 2,331 high risk patients
  – Aprotinin vs Tranexamic acid vs ε-Aminocaproic acid

• Halted early due to increased 30-day mortality: 6% aprotinin vs 4% others

• Despite decreased bleeding: 9.5% aprotinin vs 12% others

• Now withdrawn from the market
  – Even though it may help some patients (pediatrics?)
  – 21 years after initial efficacy trial (Lancet)
The circle of life

Abstract-based medicine  Have you noticed?

Small RCT  Broad Use  Large Retrospective Review

Small RCT  Large RCT

“Personalized” medicine  We need to study this

Copyright Regents of University of Michigan 2015
“Comparative Effectiveness”

• Efficacy: Can it work?
  – Under ideal circumstances

• Effectiveness: Does it work?
  – In real world patients treated in ordinary clinical settings

• Efficiency: Is it worth it?
  – Value of intervention in relation to resources it consumes
A new responsibility

The Value of Phase 4 Clinical Testing
Gus J. Vlahakes, M.D.

In Defense of Pharmacoepidemiology — Embracing the Yin and Yang of Drug Research
Jerry Avorn, M.D.

Observational Medical Outcomes Partnership
Why isn’t efficacy enough?

- The “C” in RCT: control
  - The clinical process can’t be recreated
    - Intensive insulin therapy
  - The intervention is complex
    - Perioperative beta blockade
  - The patients are not your patients
    - Coronary revascularization for preop optimization
  - The intervention requires learning
    - Videolaryngoscopy for failed airway
Why isn’t efficacy enough?

- Risk and benefit must be considered
  - The adverse events are not observed due to rarity of the complication
    - Aprotinin
    - Vioxx
  - The adverse events are not observed due to inadequate follow-up period
    - Hormone replacement therapy
- Risk and benefit are not uniform
  - Lower benefit in some patients
  - Higher risk in some patients
“Comparative Effectiveness”

• Efficacy: Can it work?
  – Under ideal circumstances

• Effectiveness: Does it work?
  – In real world patients treated in ordinary clinical settings

• Efficiency: Is it worth it?
  – Value of intervention in relation to resources it consumes
An obsolete “hierarchy” of evidence

BOX B: **Levels of Evidence**

**Category I:** Evidence from at least one properly conducted randomized controlled trial.

**Category II-1:** Evidence from well-designed controlled trials without randomization.

**Category II-2:** Evidence from well-designed cohort or case-control analytic studies, preferably from more than one center or research group.

**Category II-3:** Evidence from multiple times series with or without intervention or dramatic results in uncontrolled experiments, such as the results of the introduction of penicillin treatment in the 1940s.

**Category III:** Opinions of respected authorities, based on clinical experience, descriptive studies and case reports, or reports of expert committees.

### About as useful as...  

#### A realistic evidence based rating scale:

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 0</td>
<td>Things I believe</td>
</tr>
<tr>
<td>Class 0a</td>
<td>Things I believe despite the available data</td>
</tr>
<tr>
<td>Class 1</td>
<td>Randomised controlled clinical trials that agree with what I believe</td>
</tr>
<tr>
<td>Class 2</td>
<td>Other prospectively collected data</td>
</tr>
<tr>
<td>Class 3</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>Class 4</td>
<td>Randomised controlled clinical trials that don't agree with what I believe</td>
</tr>
<tr>
<td>Class 5</td>
<td>What you believe that I don't.</td>
</tr>
</tbody>
</table>
All RCTs are not the same

• Must evaluate
  – Background
  – Methods
  – Results
  – Discussion

• “CONsolidated Standards Of Reporting Trials” = CONSORT
• Group of Research Researchers
• Objective checklist for evaluation of an RCT
• Adopted by all major journals
<table>
<thead>
<tr>
<th>Rule and abstract</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
</tr>
<tr>
<td>Methods</td>
<td>3</td>
</tr>
<tr>
<td>Participants</td>
<td>4</td>
</tr>
<tr>
<td>Interventions</td>
<td>5</td>
</tr>
<tr>
<td>Objectives</td>
<td>6</td>
</tr>
<tr>
<td>Outcomes</td>
<td>7</td>
</tr>
<tr>
<td>Sample size</td>
<td>8</td>
</tr>
<tr>
<td>Randomization</td>
<td>9</td>
</tr>
<tr>
<td>Sequence generation</td>
<td>10</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>11</td>
</tr>
<tr>
<td>Implementation</td>
<td>12</td>
</tr>
<tr>
<td>Blinding (masking)</td>
<td>13</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>14</td>
</tr>
<tr>
<td>Results</td>
<td>15</td>
</tr>
<tr>
<td>Participant flow</td>
<td>16</td>
</tr>
<tr>
<td>Recruitment</td>
<td>17</td>
</tr>
<tr>
<td>Baseline data</td>
<td>18</td>
</tr>
<tr>
<td>Numbers analyzed</td>
<td>19</td>
</tr>
<tr>
<td>Outcomes and estimation</td>
<td>20</td>
</tr>
<tr>
<td>Ancillary analyses</td>
<td>21</td>
</tr>
<tr>
<td>Adverse events</td>
<td>22</td>
</tr>
<tr>
<td>Discussion</td>
<td>23</td>
</tr>
<tr>
<td>Interpretation</td>
<td>24</td>
</tr>
<tr>
<td>Generalizability</td>
<td>25</td>
</tr>
<tr>
<td>Overall evidence</td>
<td>26</td>
</tr>
</tbody>
</table>

- **Methods**
  - **Participants**: Eligibility criteria for participants and the settings and locations where the data were collected.
  - **Interventions**: Precise details of the interventions intended for each group and how and when they were actually administered.
  - **Objectives**: Specific objectives and hypotheses.
  - **Outcomes**: Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).
  - **Sample size**: How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.

- **Randomization**
  - **Sequence generation**: Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification).
  - **Allocation concealment**: Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.

- **Results**
  - **Participant flow**: Flow of participants through each stage (a diagram is strongly recommended).
  - **Recruitment**: Dates defining the periods of recruitment and follow-up.
  - **Baseline data**: Baseline demographic and clinical characteristics of each group.
  - **Numbers analyzed**: Number of participants (denominator) in each group included in each analysis and whether the analysis was by “intention to treat.” State the results in absolute numbers when feasible (e.g., 10 of 20, not 50%).
  - **Outcomes and estimation**: For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95% confidence interval).
  - **Ancillary analyses**: Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.
  - **Adverse events**: All important adverse events or side effects in each intervention group.

- **Discussion**
  - **Interpretation**: Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.
  - **Generalizability**: Generalizability (external validity) of the trial findings.
  - **Overall evidence**: General interpretation of the results in the context of current evidence.
Replication

• The missing link in “advancing” science
• What is it?
  – Attempts at replicating results using different datasets, different populations, or different analytic techniques
• The problems with it
  – It is boring
  – It isn’t innovative
  – We don’t promote people for copying science
• But, our patients deserve it…
Replication to Advance Science

Changes in Anesthesiology

James C. Eisenach, M.D.,

“WHAT'S new with you: "Take a look at that!" "Well that's certainly different!"

We are interested in novel and are wired to notice change in our environment. Advances in technology mean that many of us are more frequently prodded with updated information.

Prospective External Validation of a Predictive Score for Postoperative Pulmonary Complications

Valentín Mazo, M.D., Sergi Sabaté, M.D., Ph.D., Jaume Canet, M.D., Ph.D., Lluís Gallart, M.D., Ph.D., Marcelo Gama de Abreu, M.D., Ph.D., Javier Belda, M.D., Ph.D., Olivier Langeron, M.D., Ph.D., Andreas Hoef, M.D., Ph.D., Paolo Pelosi, M.D.

Abstract

Background: No externally validated risk score for postoperative pulmonary complications (PPCs) is currently available. The authors tested the generalizability of the Assess Respiratory Risk in Surgical Patients in Catalonia risk score for PPCs in a large European cohort (Prospective Evaluation of a Risk Score for postoperative pulmonary COMplications in Europe).

Methods: Sixty-three centers recruited 5,859 surgical patients receiving general, neuraxial, or plexus block anesthesia. The Assess Respiratory Risk in Surgical Patients in Catalonia factors (age, preoperative arterial oxygen saturation in air, acute respiratory infection during the previous month, preoperative anemia, upper abdominal or intrathoracic surgery, surgical duration, and emergency surgery) were recorded, along with PPC occurrence (respiratory infection or failure, bronchospasm, atelectasis, ...
Steps to replication

• Step 1: Reproducibility
  – Different analytical techniques applied to same dataset by a distinct research team

• Step 2: Replicability
  – Different random subsets of the primary dataset

• Step 3: Generalizability
  – Different datasets, patient populations, care settings
A surprisingly popular guy

Essay

Why Most Published Research Findings Are False

John P.A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and other factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9, 11] that the high

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider.
Is there an ideal technique?

RANDOMIZED, CONTROLLED TRIALS, OBSERVATIONAL STUDIES, AND THE HIERARCHY OF RESEARCH DESIGNS

JOHN CONCATO, M.D., M.P.H., NIRAV SHAH, M.D., M.P.H., AND RALPH I. HORWITZ, M.D.


(Slide Courtesy Andreas Hoeft)
The ultimate in replication

Effects of Volatile Anesthetic Choice on Hospital Length-of-stay

A Retrospective Study and a Prospective Trial

Impact of Volatile Anesthetic on Hospital Length-of-Stay

Ratio of geometric means (95% CI)

Prospective Trial:
- Isoflurane vs. Sevoflurane

Retrospective Study:
- Isoflurane vs. Sevoflurane
- Isoflurane vs. Desflurane
- Sevoflurane vs. Desflurane
What is the next aprotinin?

- Repeated examples of “irrational exhuberance”
- Which current *en vogue* clinical practice will fail us
- Anesthetic technique
  - Regional or neuraxial anesthesia?
  - Brain function monitoring?
- Fluid therapy
  - Colloids?
  - Tranexamic acid?
- DVT prophlaxis?
- Surgical site infection prophylaxis?
Expanding the toolset

- Observational datasets to be considered
  - Administrative (eg NIS, Premier, CMS, HCCI)
  - Clinical registries (eg STS, McSPI, NSQIP, AQI)
  - Electronic health records

- Statistical tools
  - Case control
  - Propensity score matching & stratification

- Data gathering
  - Objective outcomes (death, cost, laboratory values)
  - Clinical documentation
  - Understand data limitations
We have thousands of data collectors!!

Anesthesia Faculty, Residents & CRNAs
Curriculum

- Being smart might not be enough to make progress
- It can be very tedious
- It can be frustrating to read (or perform) new types of research
Achievable Curriculum

1. Outcomes research overview
2. Privacy, security, IRB, and access
3. How to design an answerable question
4. Data sources available for observational research
5. Evaluate 2 clinical questions
6. Basic statistics for non-statisticians
7. More basic statistics for non-statisticians
8. Reading / writing an observational manuscript
9. Bayesian analytics
10. Artifact reduction

• Each is 45 minutes long
• http://www.mpogresearch.org/curriculum
Curriculum

MPOG offers a wide variety of online educational materials for the perioperative research community. Feel free to download and view.

If you have a lecture you would like to contribute, please email Tory Lacca

Research Boot Camp Lectures

- Lecture One: Outcomes Research Overview
- Lecture Two: Data Sources Available for Observational Research
- Lecture Three: How to Design an Answerable Question
- Lecture Four: Basic Statistics for Non-Statisticians - Descriptive & Univariate Techniques
- Lecture Five: Evaluation and Development of a Testable Research Hypothesis
- Lecture Six: Privacy, Security, IRB and Access
- Lecture Seven: More Basic Statistics for Non-Statisticians - Multivariate Techniques
- Lecture Eight: Writing an Observational Manuscript

MPOG Instructional Videos

- Instructional Video for Research Application
- Instructional Video for Epic Mapping Utility

ASA 2013 MPOG Retreat Lecture

- MPOG ASA '13 Bayesian Analytics Lecture by Dr. Tim Houle

ASA 2014 MPOG Retreat Lecture

- MPOG ASA '14 Artifact Reduction in Electronic Databases by Dr. Leif Saager
The problem with outcomes databases

• Excuses
  – “My patients are sicker”
  – “Our cases are harder”
  – “We are unique”

• Need risk adjustment data elements
  – Preoperative comorbidities
  – Intraoperative events
  – Surgical details
  – “Systems” issues (hospital volume, staffing, etc)

• Need to control confounders – just like an RCT hopes to
“Outcome” data structures

Laboratory – postop troponin peak, creatinine values

Demographics – Date of Death

Financial – Dialysis charge

Pharmacy – diuretic

Local outcome registry

PACU LOS
Lessons learned

• Consider alternate study methods and statistical tools

• Collect all the data possible
• Collect the data you need (now and in near future)
  – As granular as possible
  – Integrate with other sources of the “truth”
  – Iterative content development

• Know your data
  – Data quality, consistency, and shortcomings
  – Look at the data yourself

• Specific hypothesis \(\rightarrow\) quality studies
Tips – The Question

• Ensure the question can’t be answered using traditional, small, randomized trials

• Ensure the question hasn’t ALREADY been answered – do a detailed literature search

• Focus on specific events / issues
  – With low incidence (ie < 2 or 3%)
  – Difficult to randomize / control
  – Adverse events that are side effects of treatments
  – Low frequency treatments

• Do a power analysis
Tips – The Data

• Know the data quality
  – Watch or participate in clinician entry of data
  – Hand review THOUSANDS of records to see what people are entering
  – Compare to some version of the truth as a check step

• Use multiple sources and combine
  – Clinical, administrative, financial

• Admit reality
  – Realize when data is NOT good enough to be the basis for a research question
Tips – The Methods

• Establish a clear, testable hypothesis

• Establish quantitative outcomes
  – If you can’t measure it, you can’t study it
  – Get rid of abstract concepts like “hemodynamic instability” or “poor compliance”

• Have the data collection, patient exclusions, statistical analysis, and study plan in place BEFORE you extract the data

• Involve the database and statistical experts early
Summary

• Establish where the data is in the “circle of life”

• Read beyond the abstract

• Assess the care process and patients in the study

• Wait for replication studies before jumping on the bandwagon

• Choose the right studies for observational research