Knowing what is so: Evidence-based medicine
Why clinical research is vital to the veterinary profession
Ernest Codman, 1869-1940: A Pioneer of Evidence-Based Medicine: The end result idea

- "The End Result Idea": Every hospital should follow every patient long enough to determine if the treatment was successful
- "If not, inquire why not?" to prevent future failures
- End Result - System of Hospital Organization: Tracking and publishing the results of hospitals and of individual surgeons
- Led to establishing the American College of Surgeons

- Result: Harvard fired him as surgery instructor
- "acerbic, bitter, flamboyant, eccentric, combative, outspoken"
- Codman's Sign, Classification, Tumor, Exercises, Triangle
- http://www.whonamedit.com/

EBM Story Timeline & Dignitaries

- Early 1900's - Ernest Codman’s "The End Result Idea" led to empirical outcome evaluation
- Late 1940's – Professor Austin Bradford Hill established the RBCT as the evidence "gold standard"
- 1972 – Dr. Archie Cochrane’s bestselling book Effectiveness and Efficiency: Random reflections on health services
  - Stated that no more than 10% of doctors’ interventions were proven to do more good than harm
  - Challenged doctors to select procedures and interventions on basis of the RBCT (randomized blinded controlled trial)

Use of randomisation in the Medical Research Council’s clinical trial of streptomycin in pulmonary tuberculosis in the 1940s

- Demand for Streptomycin, isolated in 1943, exceeded supply
- Treatment cost 2x average 1947 car cost
- Spontaneous cures confounded studies
- Poorly controlled studies led inefficacious, harmful treatments
- Blind randomization relieved MRC clinicians of deciding who would be treated

- Physician smoking study – smoking and lung cancer

EBM Story Timeline & Dignitaries

- Early 1900's - Ernest Codman’s "The End Result Idea"
- Late 1940's – Professor Austin Bradford Hill established the RBCT as the evidence "gold standard"
- 1972 – Dr. Archie Cochrane challenged physicians to base the selection of their practices on the RBCT
- Late '70s to present – Studies undertaken after Archie Cochrane’s challenge accumulated evidence of problems with:
  1. Current practices
  2. Evaluating new interventions
  3. Disseminating new information
  4. Synthesizing information
Every aspect of medical practice progress
Resulted in the Cochrane Reviews - http://www.cochrane.org/

What is the situation in veterinary medicine??

What proportion of human clinical medicine is evidence-based?

Resource Guide (Andrew Booth):

- 18 studies between 1995 and 2000 determined the strength of evidence supporting clinical procedures
- Of ~28 procedures per study (range 40 to 1,990)
  - 38% were supported by RBCT Type I (strongest) evidence
  - 22% were not supported by convincing experimental or non-experimental evidence - Dogma!

What is this % in veterinary medicine??

The ‘dogma’ problem

- Dogma - Those beliefs put forth without supporting empirical evidence with which to judge strength
  - Not presented for efficiency (textbooks, notes)
  - Unchecked hypotheses or uncritical observation without awareness of the effects of chance, biological variation and observer bias (authoritative or established opinion)
- Repetition across sources or people, whatever their qualifications, does not change the status of such information

Some dogma is right, some dogma is wrong; the problem is which is which?
The problem of anecdotal evidence

- Anecdote – A report of the occurrence of desired event, such as medical recovery
  - Case reports
- Often interpreted as due to therapy applied and thus validating theory on which therapy was based
- Problems:
  - Probability of apparently unusual events is higher than expected by intuition (birthday paradox ~ 30:70%)
  - Unrecognized factors (confounders) may invalidate initial prediction (e.g., dog wasn’t as sick as believed or the condition was misdiagnosed)
- Because anecdotes are extremely weak evidence, accumulating similar anecdotes does not significantly increase support for a theory
- Case series

The plural of ‘anecdote’ is not ‘evidence’!

Consider information learned in the DVM curriculum:

- Most of the veterinary curriculum is taught as dogma
  - Neither is the information nor the links to the information needed to judge credibility provided
  - Lectures, core textbooks
  - Habits of continual deliberate, systematic, critical thinking are not well developed (and, given resource limitations, possibly cannot be)
- Analogical evidence is the most prevalent basis of reasoning taught in much of the curriculum
- Without formal followup mechanisms, most clinicians’ individual case experience remains anecdotal

Resource Efficiency vs. Educational Efficacy

The necessity but weakness of analogical evidence

- Analogical Evidence - Reasoning from comparing known similarities between two systems in which a relationship known to exist in one is reasoned to exist in the other
  - If drug X has been shown to be effective against disease Y in a species Z, reasoning that the same relationship exists between similar drugs, similar diseases or similar species

Drug X Species Z
Drug Y Known Species U
Disease Y

- Susceptible to error because of the likelihood that different but unknown factors are operating in the two systems
  - Very common basis for reasoning in veterinary medicine
  - Necessary basis for action when empirical evidence is lacking
  - Particularly true for minor species
  - Mechanisms are often established in convenient species (rodents) and then extrapolated to species in which direct investigation is impractical

Considering Empirical Evidence

- Because the opportunity to verify (repeatability) and to assess evidence strength is key, the methods used to acquire the evidence must be described or referenced sufficiently to enable this verification and assessment (transparency)
  - When authors don’t provide key study design information, because of hard-wired cognitive biases they are unlikely to have executed a strong study
- Remember: The empirical evidence can be correct (e.g., the sun “rises” predictably) but the underlying theory that it is believed to support is wrong (e.g., the sun orbits around the earth)

Evidence of problems with common current practice

Tonsillectomy Study (Pseudodoxia Pediatrica NEJM 232:691)

- 389 11 year old children with tonsillitis
- 234 (55%) Tonsillectomy not Recommended
- 174 (45%) Tonsillectomy Recommended
- Examined by physicians
- Examined by other MD’s

(Blind to previous exam results)

What do you think the other MD’s recommended?
Clinician-Discoverers – Marshall, Warren, and H. pylori
NEJM 353(23):2421-2423

The Nobel Prize in Medicine 2005
"For their discovery of the bacterium Helicobacter pylori and its role in gastritis and peptic ulcer disease"

Fascinating story of:
- How they challenged dogma that:
  - Due to acid, the stomach was sterile
  - Stress and spices cause ulcers
  - Serendipity
  - How advancing technology enabled their research
- How much the medical community resisted their findings

Dr. Eddy’s JAMA essay on medical decision problems
Eddy DM. The Challenge. JAMA 263:287-290

Summarized findings similar to the earlier Tonsillectomy study:
- Chance of tonsil removal was 8% in one community, 70% in a neighboring community
- Given written descriptions of surgical problems, 50% of surgeons recommended surgery, 50% not
  - When surveyed two years later, 40% changed their recommendations

"Physicians are in the impossible position of facing uncertain outcomes from different actions, but having to act anyway"
"Variability occurs because physicians must make decisions about problems in complex systems under difficult circumstances with little support"

How well are veterinarians dealing with even less support doing?"
Do medical errors remain a big deal?

Investigator enthusiasm is inverse to degree of control

Table 3. Degree of control versus degree of investigator enthusiasm for portacaval shunt operation in 51 studies with at least ten patients. The table is revised from Grace, Muench, and Chalmers (8), table 5, p. 685.

<table>
<thead>
<tr>
<th>Degree of control</th>
<th>Marked</th>
<th>Moderate</th>
<th>None</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well controlled</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Poorly controlled</td>
<td>10</td>
<td>2</td>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>24</td>
<td>7</td>
<td>10</td>
<td>41</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>13</td>
<td>63</td>
<td>50</td>
</tr>
</tbody>
</table>

Fisher’s Principles of DOE:
• Concurrent comparisons
• Randomization
• Replication

Problem with Information Synthesis & Diffusion
A comparison of results of meta-analysis or randomized control trials and recommendations of clinical experts: Treatments for Myocardial Infarction

- Contrasted results from 182 RBCTs with recommendations of 43 review articles and 100 textbook chapters by publication date
- Classified expert recommendations as:
  1. Routine – use routinely
  2. Specific – used only in selected patients with particular indication
  3. Rare/Never
  4. Experimental
  5. Not mentioned

This study took a huge amount of work! And it had big impact – 895 citations to date

Continuing to recommend a questionable therapy on basis of biological plausibility

RCT meta-analysis vs. Expert Recommendations

Problem of both information synthesis and diffusion
Thorax pump failure shows weakness of biological plausibility

Intended to improve blood flow during acute cardiac failure resuscitation
• (++) Biological plausibility
• (++) Lab experimental results
• But equivocal field results

Relying on experts is dangerous

Sir William Arbuthnot Lane
Scottish surgeon, 1856 - 1943
• Developed internal fixation to align fractures
• Used silver wire, then screws followed by plates and screws
• Hence, the label "Lane's plates"

He also promoted erroneous ideas

Sir Lane:
• Was "eccentric", regarding humans as machines
• Performed many total colectomies as a cure for "auto-intoxication"
• 10% mortality risk

Systematic Reviews: Synthesis of the best evidence for clinical decisions
Ann Intern Med 126:376-380

Systematic review procedures evolved to correct problems of the narrative reviews

<table>
<thead>
<tr>
<th>Feature</th>
<th>Narrative Review</th>
<th>Systematic Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Often broad in scope</td>
<td>Often a focused clinical question</td>
</tr>
<tr>
<td>Sources and search</td>
<td>Not usually specified, potentially biased</td>
<td>Comprehensive sources and explicit search strategy</td>
</tr>
<tr>
<td>Selection</td>
<td>Not usually specified, potentially optional</td>
<td>Criterion-based selection, using explicit applied</td>
</tr>
<tr>
<td>Appraisal</td>
<td>Variable</td>
<td>Rigorous critical appraisal</td>
</tr>
<tr>
<td>Synthesis</td>
<td>Often a qualitative summary</td>
<td>Quantitative summary*</td>
</tr>
<tr>
<td>Inferences</td>
<td>Sometimes evidence-based</td>
<td>Usually evidence-based</td>
</tr>
</tbody>
</table>

Reporting standards are rapidly evolving
ARRIVE, BEME, CONSORT, MOOSE, REFLECT, STARD
http://www.equator-network.org/

What's going on here?
Our minds function roughly in two parts, conscious and subconscious.

System I: Thinking fast
- The massive Elephant
- Automatic, effortless, no sense of thought control
- Dominates under stress, alarm, anger
- Habitual, trained by repetition
- Emotional, intuition, "gut feeling"
- System I: Thinking fast

System II: Thinking slow
- The small Rider
- Deliberate, effortful, requires concentration and energy investment
- Capacity exhausted under stress
- Abstract
- Rational, reasoning, logical
- Explicit learning
- Controlled, often quenching impulses
- Danger of Inattention, distraction
- The 'pattern matching' Expert

The power of System 1 (fast thinking) is amazing

Most understand this quickly:

Aoccdng to a rschearch at Cmabrigde Uinervtsiy, it deosn't mttær in what orde the itteers in a wrod are, the olny iprmoen thing is taht the frist and lsat itteer be at the rght plc ae. The rset can be a toalt mse and you can sittl reed it woutht porbleem. Ttis is bcsue the huamn mnid deos not raed ervey itteer by istlef, but the wrod as a wloe.

Given above and SpellCheck, one might wonder why we invest so much time in learning to spell!

But this also gives clues to why it is easy to misread things such as labels!

There is more to this meme than is represented above - Google 'Davis Cambridge jumbled letters'

System I: Thinking fast | System II: Thinking slow
---|---
The massive Elephant | The small Rider
Automatic, effortless, no sense of thought control | Deliberate, effortful, requires concentration and energy investment
Dominates under stress, alarm, anger | Capacity exhausted under stress
Habitual, trained by repetition | Abstract
Emotional, intuition, "gut feeling" | Rational, reasoning, logical
System I: Thinking fast | Explicit learning
Impulsive, instinctual behavior | Controlled, often quenching impulses
Danger of Inattention, distraction | The 'pattern matching' Expert
The 'pattern matching' Expert | The Novice Student

Optical illusions illustrate the 'education role' of System II and the dominance of System I

“My Wife and My Mother-in-Law”
WE Hill, Puck 6/11/1915
(also called the Boring Figure)
- Is it wife's Ear or mother-in-law's Eye?
- Is it wife's Chin or mother-in-law's Nose?
- Is it wife's Necklace or mother-in-law's Mouth
- The Point: Once System II educates System I as to what you see, System I makes it hard for System II to see something else

Human reasoning like playing pool on a table with a warped (biased) slate

- Nature wired biases into our brains
- Evidence shows that knowledge of these biases does not significantly reduce their effect
- Processes must be designed to minimize the opportunity for each
- Example: collective application of the scientific method and process
- Published RBCT's

Takeoff on Steven Pinker’s The Blank Slate: The modern denial of human nature (2002)

If you remember nothing else from today, remember

Multiple cognitive biases affect our reasoning

Cognitive biases associated with medical decisions: a systematic review
BMC Medical Informatics and Decision Making (2016) 16:138

The big ones affecting veterinary decision-making likely include:
- Anchoring effect
- Availability bias
- Confirmation bias
- Framing effect
- Hindsight bias
- Lead-time bias
- Overconfidence bias
- Recency effect
- Search satisfying bias
- Survivorship bias

The Big Question: Why do we have these?
Error Mgmt Theory: Nature biased us toward the least costly reasoning error

<table>
<thead>
<tr>
<th>Snake Avoidance</th>
<th>True World State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed World State</td>
<td>Snakes Present</td>
</tr>
<tr>
<td>蛇</td>
<td>Correct Detection</td>
</tr>
<tr>
<td>蛇不存在</td>
<td>False Negative (costly error)</td>
</tr>
</tbody>
</table>


We have incredible, but flawed, sense-making capacity

“We try to make sense of the world, and in our brain nature has gifted us with a double-edge sword”

• “It is a complex and sophisticated information processing machine, capable of understanding not only the universe itself but of understanding the process of understanding”

• “By the same process of forming beliefs about how the universe works, we are most capable of self-deception and illusion, of fooling ourselves while we try to avoid being fooled by nature”

The first principle that you must not fool yourself—and you are the easiest person to fool.

Richard Feynman, 1974 CalTech Commencement

This dual-process reasoning model is being incorporated into clinical teaching to reduce cognitive bias

Interactionist Theory of Reasoning: Opportunity

New interactionist theory of reasoning has considerably more explanatory power than older theories:

• Explains why confirmation bias occurs in our reasoning

• Explains why we are more efficient at evaluating arguments than producing them

• Explains why in a social exchange our responses tend to be stronger than our statements

• Supports the wisdom of the complete scientific process and of the adversarial judicial process

• Supports working in pairs that trust but challenge each other

The Question: How do we restructure current processes to take advantage of our natural reasoning structure?
Interactionist Theory vs. Intellectualist Theory of Reasoning:

Conventional intellectualist view of reasoning:
- Reasoning evolved as a means to improve individual cognition for arriving at better beliefs and decisions for ourselves by ourselves
- We are biased toward finding reasons supporting our view point because:
  - Goal is to justify our actions and to convince others to share our beliefs
  - Presenting reasons that undermine this support is counterproductive
- We are more demanding and objective in evaluating others’ reasons than in producing our own to:
  - Avoid being deceived by poor or fallacious arguments into accepting false ideas
  - Be justified in revising our beliefs when presented with sound reasons for change

Emerging interactionist view of reasoning:
- Reasoning evolved primarily as a tool to facilitate social interaction
- We are biased toward finding reasons supporting our view point because:
  - Goal is to justify our actions and to convince others to share our beliefs
  - Presenting reasons that undermine this support is counterproductive
- We are more demanding and objective in evaluating others’ reasons than in producing our own to:
  - Avoid being deceived by poor or fallacious arguments into accepting false ideas
  - Be justified in revising our beliefs when presented with sound reasons for change

Illusion of Understanding (Fluency): Why can’t we explain as much or to the depth that we think we can?

- Intuition provides us simplified, coarse, often good enough analyses, causing “illusions of understanding”
- “Illusion of fluency” arises when we confuse our familiarity or recognition with understanding
- We overlook complexity by failing to recognize it or by over-simplifying it
- We fail to distinguish between readily accessible knowledge (or in the heads of others) and that which is actually in our heads

“My success is ... due to how I deal with not knowing ... it’s dealing with what one doesn’t know that’s more effective than knowing.” Ray Dalio, Bridgewater Associates, 2015

Evidence strength classification is a pyramid from weak to strong

- **Class I**: Systematic reviews based on multiple RBCT’s
- **Class II**: High quality clinical trials using historical controls, sound observational studies
- **Class III**: Uncontrolled case series
- **Class IV**: Expert opinion or extrapolated from other studies

Work at uncovering our IOED’s (Illusions of Explanatory Depth)

IOED occurs when:
- Our mental model of the process is too simple, too abstract, or flat
- We don’t understand the causal linkages well enough to describe how all of something works

To uncover one’s IOED for a complex problem, explain the causal chain to a depth of 5 why’s or so, noting one’s knowledge gaps
- Teach it to a group of sharp students
- Invest 10 minutes drawing out the causal web in front of a peer, encouraging them to critique it
- Invest 30 minutes drawing out key causal linkages and then verify with a good textbook

Lack of quality evidence is a major problem for veterinary medicine

Aragon CL, Budsgberg SC. Applications of evidence-based medicine: cranial cruciate ligament injury repair in the dog.

- **OBJECTIVE**: Evaluate the literature reporting surgical interventions for canine cranial cruciate ligament (CCL) injury using EBVM
- **STUDY DESIGN**: Systematic literature review
- **METHODS**: Search through Medline, PubMed, VIN, and CAB Abstracts performed August 2004 identified 240 sources of information
- Studies were evaluated for study design (retrospective, prospective, randomization), surgical technique, short- and long-term follow-up, and evidence classification

Evaluating the evidence strength:

- **Strong**: Cochrane Systematic Reviews Other RCT & Meta-Analysis
- **Moderate**: Evidence Guidelines Evidence Surveys Evidence Summaries RCT’s Case Controls, Control Studies Clinical Research Case Studies Other Reviews of the Literature
- **Weak**: Case Reports, Case Series, Practice Guidelines, etc.


- **RESULTS**: 28 (12%) sources qualified for evidence classification
- **No class I or class II studies** were present, 5 studies were class III and 23 studies were class IV
- **17 studies** were retrospective and 11 were prospective
- Results ranged from:
  - Variety of subjective findings including clinical impression, radiographic analysis, synovial fluid analysis, gross pathology, and histopathology
  - Objective results, although infrequent, included force plate analysis and cadaveric biomechanical testing

The human literature is sufficiently deep that they seldom use studies less than class II in systematic reviews
Poor evidence strength results for most common veterinary surgical procedure

- Class I: systematic reviews based on multiple RCTs
- Class II: high quality clinical trials using historical controls, sound observational studies
- Class III: uncontrolled case series
- Class IV: expert opinion or extrapolated from other studies


- CONCLUSIONS: At this time, the application of EBM suggests that there is not a single surgical procedure that has enough data to recommend that it can consistently return dogs to normal function
- Assessing and categorizing the evidence becomes increasingly important as more data becomes available and research quality improves
- CLINICAL RELEVANCE: An EBM paradigm did not provide sufficient evidence favoring 1 surgical technique for management of canine CCL injury

Hofmeister, King, Read, Budberg. Sample size and statistical power in the small-animal analgesia literature

- Searched Pubmed and selected journals between 1994 and 2004 for analgesia papers that reported:
  - 'no difference' between experimental groups in the abstract, results or conclusion
  - Were randomized, prospective and blinded
- Found 46 studies of which 22 provided sufficient data for power analysis
- Performed power analyses to determine sample size for detection of 20%, 50% and 80% treatment effect with power of 80%
  - 5 (23%) had sufficient power to detect a 20% treatment effect
  - 12 (54%) had sufficient power to detect a 50% treatment effect
  - 18 (82%) had sufficient power to detect an 80% treatment effect

Perino LJ, Hunsaker BD. A review of bovine respiratory disease vaccine field efficacy

From some 800+ reports identified:
- "We are impressed with the small number of useful reports... while there are hundreds of reports in the literature, most suffer from one or more design flaws or limitations... it suggests that we may be making less than optimal recommendations on vaccine use because of a lack of clinically relevant information."
- "It is time to critically evaluate vaccination as a management tool"

Keene BW. Towards evidence-based veterinary medicine

- During a period in which 60 studies relevant to a particular veterinary clinical question were published in the veterinary literature, 5,400 studies were published in the human literature for the equivalent human clinical question
- Almost a 100 to 1 ratio
- Many of these were based on inherently stronger study designs
- Implies that we can not follow the human EBM path of combining existing evidence but that we have to generate strong evidence first

A way forward?

1. Establish the structure to manage information exchange and translation
2. Assemble species x discipline interest groups and establish the communication structure within
3. Establish criteria for determining the important clinical questions
4. Determine the most important clinical questions
5. Perform systematic reviews on each to establish the state of evidence and identify the weak but critical evidence
6. Establish multi-group collaborative studies to generate strong evidence
7. Perform Cochrane-style systematic reviews
8. Update the systematic reviews as further evidence becomes available

I don't have all the answers but I believe that for the good of the profession the discussion must start. Our clients do (or soon will) expect nothing less!
The Major Paradigm Shift

- **Classical Belief** – Cognitive errors were due to process mistakes committed by a perfectly reasoning machine
- **Emerging Understanding** – Cognitive errors are the consequences of the intrinsic flaws of an imperfectly reasoning machine

**Final Takeaway:** Understanding cognitive bias doesn’t reduce it! The only thing that does is following procedures designed to counter it!