Knowing what isn’t so: Evidence-based medicine

Why clinical research is vital to the veterinary profession

What is “Evidence-based Medicine” (EBM)?

EBM: "an approach to practice in which the clinician is aware of the evidence in support of their clinical practice and of the strength of that evidence" (1992, McMaster U, Hamilton, Ontario)

Practicing EBM is simply practicing clinical medicine knowing the answers to two critical questions:

1. What is the best evidence for that? “that”: diagnostic test, prognosis, therapy, preventive measure
2. How good (strong) is this evidence?

Implies clinicians are aware of evidence types, are capable of finding that evidence, and possess skills to critically assess it

What is “Empirical Evidence” (Facts)?

- Evidence: Knowledge obtained by looking rather than reasoning or feeling
- Empirical: based on verifiable observation or experiment, not on theory or pure logic (OED)
- Strengthened by rigorous observations or experiments unequivocally supporting or refuting a hypothesis
- Weakened by the opportunity for other explanations to account for the findings
  - The reader’s judgment as to the degree of this opportunity
- Stronger Study – Less opportunity for alternative explanations
  - RBCT – Randomized, blinded, concurrent controlled trial
  - Weaker Study – Greater likelihood that alternative explanation holds
  - Even intrinsically strong study designs can be seriously weakened by execution errors (detail, details, details, . . .)

Every little kid is an empiricist: “Don’t touch the stove, it’s hot!”

EBM resulted from coalescing threads:

- Teaching of Literature Evaluation
- Emergence of Clinical Epidemiology
- Historical “Tree Shakers”
  - “Evidence-based medicine” first used
  - 1992
  - ~130k Medline hits total
  - ~12,500 EBM papers in 2015
  - Increasing trend at ~13% / yr
- Evolving Computer Technology
- New McMaster PBL Medical Curriculum
- Evidence-based Veterinary Medicine?
  - ~750 Medline hits total (0.6%)
  - ~65 papers /yr (0.5%)
  - Flat - no trend

EBM Story Timeline & Dignitaries

- Early 1900’s - Ernest Codman’s “The End Result Idea” led to empirical treatment outcome evaluation
  - Conflict (“tree shaking”) is important to progress
  - Early “Tree Shakers”:
    - Florence Nightingale
    - Giuseppe Semmelweis
    - Joseph Lister
    - John Snow
    - William Harvey
    - William Osler
    - Abraham Flexner
  - “Don’t touch the stove, it’s hot!”

The road disappearing into uncertainty

Evidence-based Medicine (EBM)

- What is it?
- Why is it important?
- What can Veterinary Medicine learn from the experience of EBM in Human Medicine?
  - Pitfalls?
  - Shortcuts?

Most discussions are how to do EBM (handouts);
My purpose is providing why to do EBM!
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

Only recently adopted as major part of AHCA!

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 2.4 average 1947 car cost
- Spontaneous cures confounded studies
- Poorly controlled studies led ineffectual, harmful treatments
- Blind randomization relieved MRC clinicians of deciding who would be treated

- Physician smoking study – smoking and lung cancer
  http://www.edwardtufte.com/tufte/hill

RBCT Innovation, Major advance in medical science-methods!

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)

What proportion of clinical medicine is evidence-based?

Resource Guide (Andrew Booth):
- 18 studies between 1995 and 2000 determined the strength of evidence supporting clinical procedures
- Of ~128 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??

What is "Dogma"?

- Those beliefs put forth without supporting empirical evidence with which to judge strength
  - Not presented for efficiency (textbooks, notes)
  - Unverified 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?

PLEASE CUB YOUR DOGMA
What is Anecdotal Evidence?
- 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’!

What is 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 X Species U

Disease Y Disease Y

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

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
- Most clinicians individual case experience remains anecdotal

Resource Efficiency vs. Educational Efficacy
- balance

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)

The importance of the “materials and methods” fine print

Evidence of problems with common current practice
Tonsillectomy Study (NEJM 232:691)

389 11 year old children with tonsillitis

Examined by physicians

214 (55%) Tonsillectomy not Recommended
174 (45%) Tonsillectomy Recommended

Sneaky Epidemiologists

Examined by other MD’s

(Blind to previous exam results) What do you think the other MD’s recommended?

SX

Doing it Again

214 (55%) Tonsillectomy Not Recommended

Examined by other MD’s

115 (54%) Tonsillectomy Not Recommended
99 (46%) Tonsillectomy Recommended

Epidemiologists couldn’t believe results

(Blind to previous exam results) What do you think the next MD’s recommended?
Evidence of Continuing Problems

Eddy DM. The Challenge. JAMA 263:287-290

Similar to the earlier Tonsillectomy study:

- The chance of having one’s tonsils removed were 8% in one community and 70% in another
- Surgeons given written descriptions of surgical problems split down the middle...half recommending surgery, half not
- When surveyed again two years later, the same surgeons often disagreed with their previous opinions, with 40% changing their recommendations
- Variability occurs because physicians must make decisions about problems in complex systems under difficult circumstances with little support
- Physicians are in the impossible position of facing uncertain outcomes from different actions, but having to act anyway

Evidence of Process Problems

Eddy DM. The Challenge. JAMA 263:287-290

- Cardiologists evaluating high quality angiograms for stenosis of coronary vessels:
  - asked to estimate whether the percentage of stenosis...was greater or less than 50%, they disagreed on 60% of the patients
  - on two successive readings of the same angiograms, the observers changed their minds from 8% to 37% of the time, depending on the vessel segment
- In general, observers looking at the same thing will disagree with each other or even with themselves from 10% to 50% of the time
- Pathologists reading human biopsy slides (Not just a clinician’s problem)
  - Expert pathologists disagreed two or more times on the interpretation of 38% of human skin melanoma biopsy specimens as benign or malignant. (Hum Pathol! 27:528-31)

Admittedly, the test cases are likely the tough ones rather than the “slam dunks”

Autopsy evidence of Physician Diagnostic Error Rates

Changes in rates of autopsy-detected diagnostic errors over time: A systematic review JAMA 289(21):2849-2856

- 53 autopsy series published 1966 - 2002
- 24% major error rate (4.1% - 49.8%)
  - Involved 1st cause of death but did not affect outcome
- 9% class I error rate (0% - 20.7%)
  - Likely resulted in death
  - Accounting for steady improvements, current major error rate is likely 8% - 24% and class I rate is likely 4% - 7%
- Of 85,000 individuals dying in US hospitals each year, without misdiagnoses 34,850 would have survived to discharge
  - Advancing technology alone is not the answer

Are Problems with Current Practices Important?

Do medical errors remain a big deal?

BMJ 2016;353:i2139 doi: 10.1136/bmj.i2139

What is it in veterinary medicine??

If we haven’t looked “well”, we don’t know!

Coin Flip!

What like this is happening in animal health?


The Nobel Prize in Medicine 2005

“For their discovery of the bacterium Helicobacter pylori and its role in gastritis and peptic ulcer disease”
Investigator enthusiasm is inverse to degree of control

Table 3. Degree of control versus degree of investigator enthusiasm for percutaneous shunt operation in 53 studies with at least ten patients. The table is revised from Grace, Muench, and Halpin (49). Table 2, p. 685

<table>
<thead>
<tr>
<th>Degree of control</th>
<th>Degree of enthusiasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well controlled</td>
<td>Marked</td>
</tr>
<tr>
<td>Poorly controlled</td>
<td>Moderate</td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>None</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

- Muench’s statistical law: “Nothing improves the performance of an innovation as much as the lack of controls”
- The rows for “poorly controlled” and “uncontrolled” studies suggest that repeated, weakly controlled trials are likely to agree and build up an illusion of strong evidence because of the large count of favorable studies
- Tables for other therapies show similar results; one must be cautious in accepting results of weakly controlled investigations

Meta-analyses of key RBCT method strength vs. effect estimates
- Effect estimates from weaker studies were larger than those from comparable stronger studies in 10 of 11 meta-analyses
- Odds Ratios from trials with inadequate blinding were 30% more “beneficial” than those with adequate blinding

Not recommending a beneficial therapy

RCT meta-analysis vs. Expert Recommendations

Continuing to recommend a questionable therapy on basis of biological plausibility

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

Problems with Information Synthesis & Dissemination

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 – 395 citations to date

Continued
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

Problems with Information Synthesis

The Medical Review Article: State of the science
- 34% met 3 of 8 criteria
- 64% met 4 or 5 criteria
- 2% met 6 of 8 criteria
- 6% used quantitative synthesis
- 42% mentioned future research directions
- “current medical reviews do not use scientific methods to identify, assess, and synthesize information”
- “Medical reviews are often subjective, scientifically unsound and inefficient”
- These methods ... are proposed to improve the quality of future reviews

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

Evolution hardwired cognitive biases into our thinking

"Human nature" is to:
- Weight information consistent with current belief heavier
- Overly commit and to ignore or discount discordant information
- Search for belief-confirming data rather than belief-refuting data
- Be overly optimistic about outcomes
- Recall the most recent, the most notable, ... 
- Stop searching prematurely

Prior belief biases observation by subtly changing perception, particularly of the vague or ambiguous
- Much of medicine involves interpretation of vague patterns (radiology, histopathology, ...) 

Critical Lessons:
- Cognitive biases occur unconsciously and despite the observer’s best intentions
- Cognitively-biased thinking is detectable by external observers but not by the thinker
- Knowledge of cognitive biases is not protective
- To avoid cognitive biases, we have to follow specific procedures
- This is why the scientific method, strong experimental design, and evidence-based medicine have developed

I may well be sawing the limb off but:

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

These proportions represent the cognitive processing balance.

The two cognitive systems differ in speed, function, and learning retention.

<table>
<thead>
<tr>
<th>System I: Thinking fast</th>
<th>System II: Thinking slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>The massive Elephant</td>
<td>The small Rider</td>
</tr>
<tr>
<td>Adaptive unconscious</td>
<td>Conscious self</td>
</tr>
<tr>
<td>Automatic, effortless,</td>
<td>Deliberate, effortful,</td>
</tr>
<tr>
<td>no sense of voluntary</td>
<td>requires concentration</td>
</tr>
<tr>
<td>control</td>
<td>and energy investment</td>
</tr>
<tr>
<td>Becomes dominate under</td>
<td>Capacity exhausted under</td>
</tr>
<tr>
<td>stress</td>
<td>stress</td>
</tr>
<tr>
<td>Habitual</td>
<td>Abstract</td>
</tr>
<tr>
<td>Emotional, intuition,</td>
<td>Rational, reasoning,</td>
</tr>
<tr>
<td>“gut feeling”</td>
<td>logical</td>
</tr>
<tr>
<td>Implicit learning</td>
<td>Explicit learning</td>
</tr>
<tr>
<td>Impulsive, instinctual</td>
<td>Controlled, often</td>
</tr>
<tr>
<td>behavior</td>
<td>quenching impulses</td>
</tr>
<tr>
<td>Danger of Complacency</td>
<td>Danger of Inattention,</td>
</tr>
<tr>
<td></td>
<td>distraction</td>
</tr>
<tr>
<td>The Expert Teacher</td>
<td>The Novice Student</td>
</tr>
</tbody>
</table>

Everyone’s first answers always come from subconscious, intuitive processing.

Our biased senses interact with our believing conscious brain through our unconscious brain.

- “We form our beliefs for subjective, personal, emotional, and psychological reasons in the environments created by family, friends, colleagues, culture, and society”
- “We defend, justify, and rationalize these with intellectual reasons, cogent arguments, and rational explanations”

Beliefs naturally come first, explanations for beliefs naturally follow.

Optical illusions provide non-trivial examples of that sensory bias.

When reasoning error costs are asymmetrical, we are biased to make the less costly error.

Concentrate steadily on the white spot for 30 seconds

When image changes, what do you see?


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 are trying to avoid being fooled by nature”

Recent lay press books cover the new brain paradigm emerging to replace of that of rational perfection

Many such books are by researchers addressing their colleagues

Should you believe it possible to be unbiased and completely rational, read one of the following:

Should you believe you are without bias, you can prove so with these on-line tests

Project Implicit
https://implicit.harvard.edu/implicit/
- “It is well known that people don’t always ‘speak their minds’, and it is suspected that people don’t always ‘know their minds’”
- “This web site presents a method that demonstrates the conscious-unconscious divergences much more convincingly than has been possible with previous methods”
- “This new method is called the Implicit Association Test, or IAT for short”
Lack of quality evidence is a major problem for veterinary medicine

Aragon CL, Budsberg 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

---

Evidence strength classification is a pyramid from weak to strong

- **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

---

Poor evidence strength results for most common veterinary surgical procedure

- **Class I:** 0 systematic reviews based on multiple RCTs
- **Class II:** 0 high quality clinical trials using historical controls, sound observational studies
- **Class III:** 5 uncontrolled case series
- **Class IV:** 23 expert opinion or extrapolated from other studies

---

The human literature is sufficiently deep that they seldom use studies less than class II in systematic reviews
Aragon CL, Budsberg SC. Applications of evidence-based medicine: cranial cruciate ligament injury repair in the dog. 

- **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.

The lack of quality evidence is a consistent finding across EBVM systematic reviews.

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."

Limited evidence in the literature almost a decade later that this challenge was taken up:
What is the long term impact of these open questions on producer usage of veterinarians?

Hofmeister, King, Read, Budsberg. 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

Most important question:
What is the minimum treatment effect difference that would lead the selection of one over the other?

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 cannot follow the human EBM path of combining existing evidence but that we have to generate strong evidence first.

The most significant contribution of veterinary systematic reviews for some time to come will be identifying critical clinical research needs:
[Also noted by Cockcroft and Holmes]

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!