NIH/NCATS Machine Intelligence in Healthcare

HOW CAN WE TRUST INTELLIGENCE (HUMAN OR MACHINE) FOR GUIDING HEALTHCARE DECISIONS?

Brian S. Alper, MD, MSPH, FAAFP
Founder of DynaMed
Vice President of Innovations and EBM Development, EBSCO Health
July 12, 2019

Presenter and Disclosures

Brian S. Alper, MD, MSPH, FAAFP

Board certifications: Family Medicine, Clinical Informatics

Founder of DynaMed

Vice President of Innovations and EBM Development, EBSCO Health (for-profit company)

Project Lead, EBMonFHIR

Key Contributor – AHRQ ACTS, PC CDS Learning Network, MCBK

Member – AAFP, ACP, AMDIS, AMIA, GRADE Working Group, G-I-N, HIMSS, HL7, ISDM, ISEHC

Our Goal in Medicine

Provide the best care…

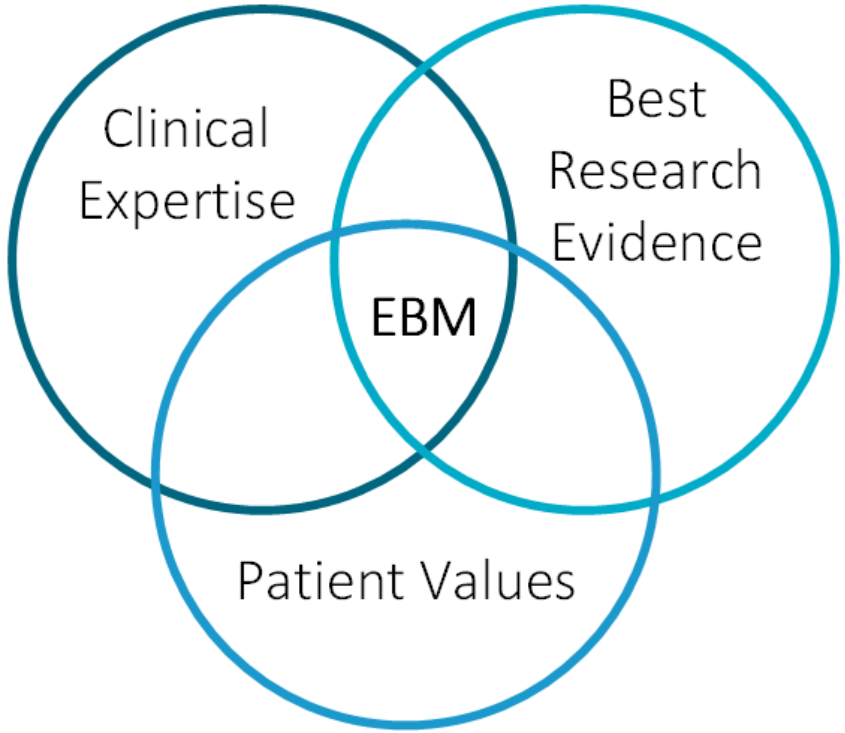
Provide the best information to guide health care decisions…

Improve health outcomes…..

…. Based on the “truth” – separating medical knowledge from folklore

Evidence Based Medicine

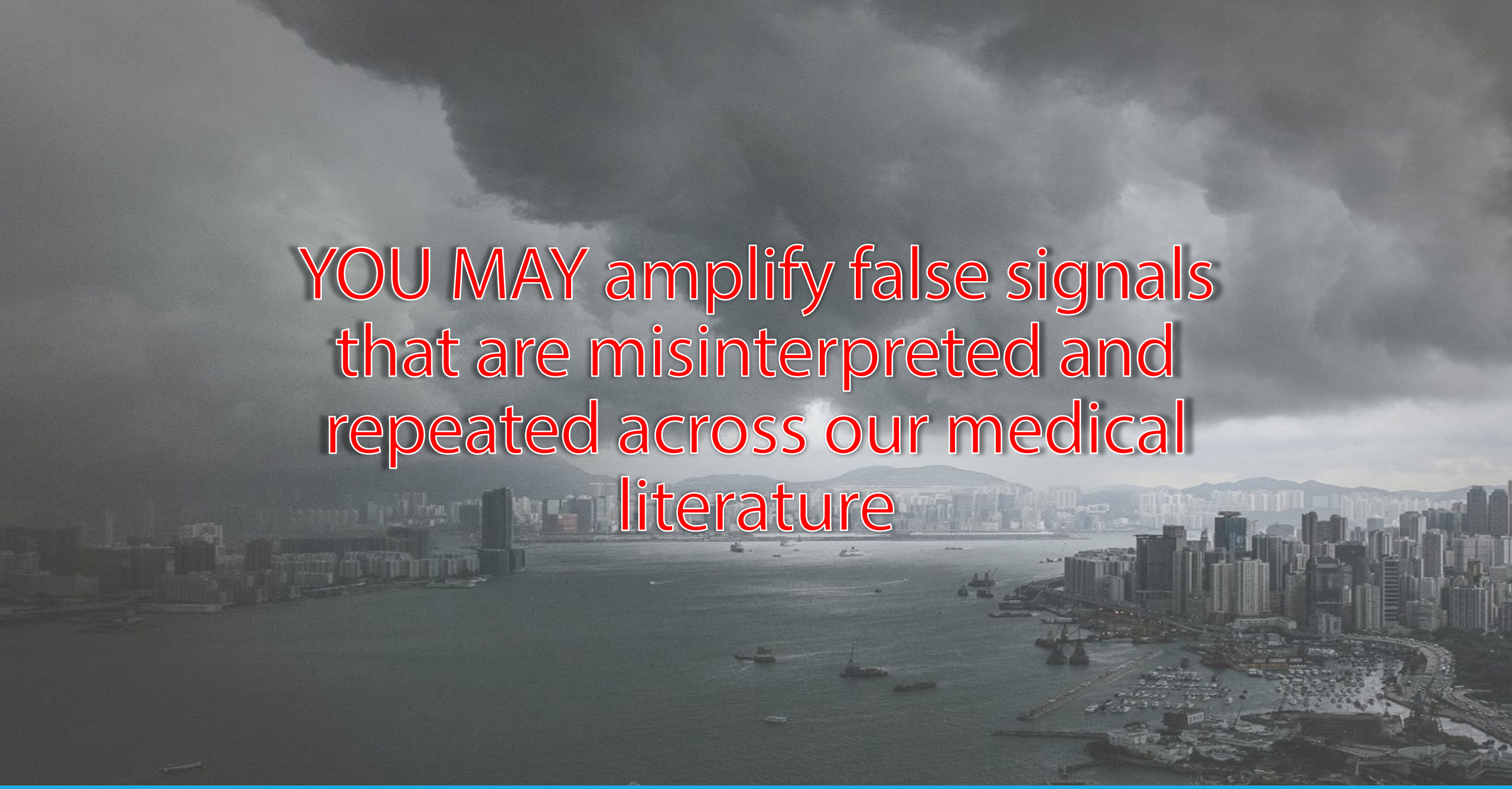
Definition: Integration of best research evidence with clinical expertise and patient values.

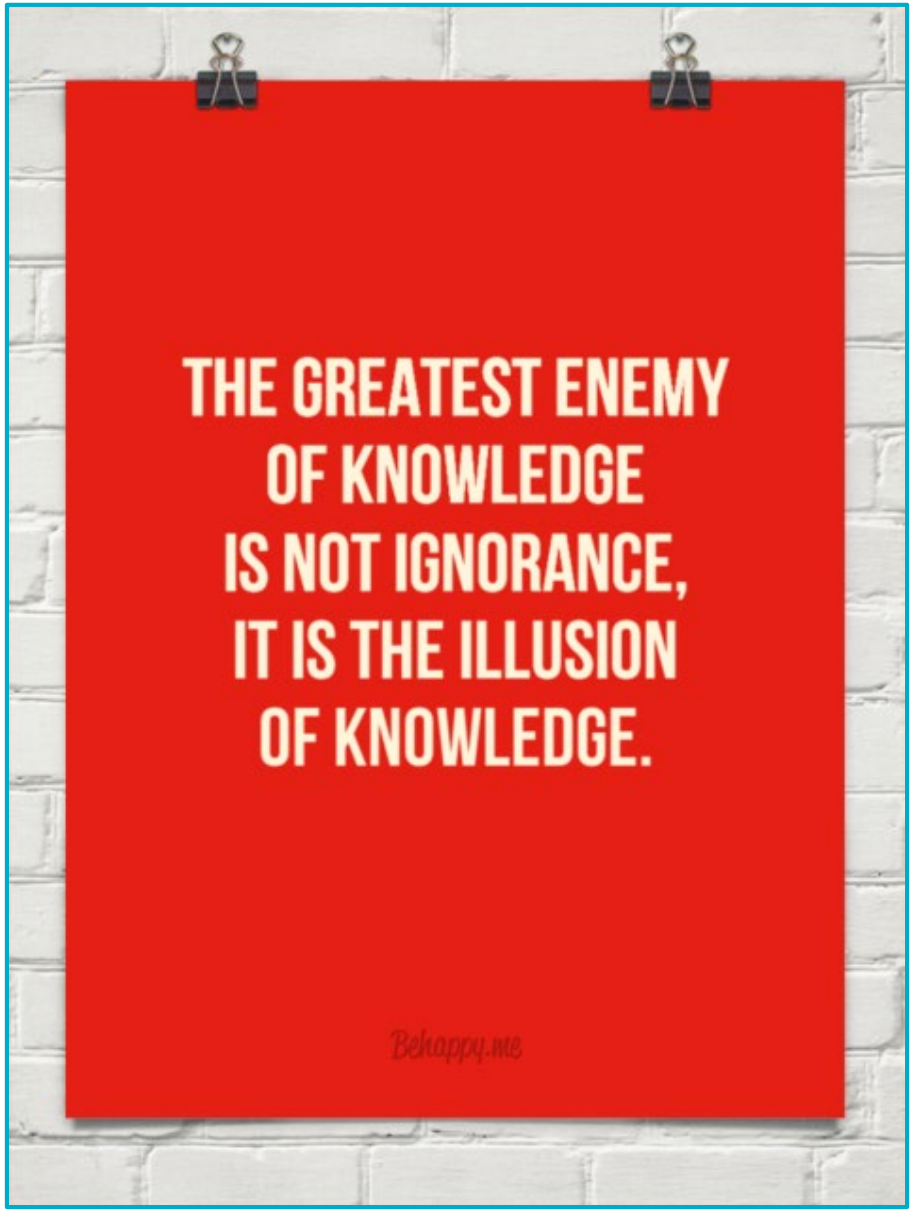


Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine. How to Practice and Teach EBM*. 2nd ed. London: Harcourt Publishers Ltd. 2000. p. 1.

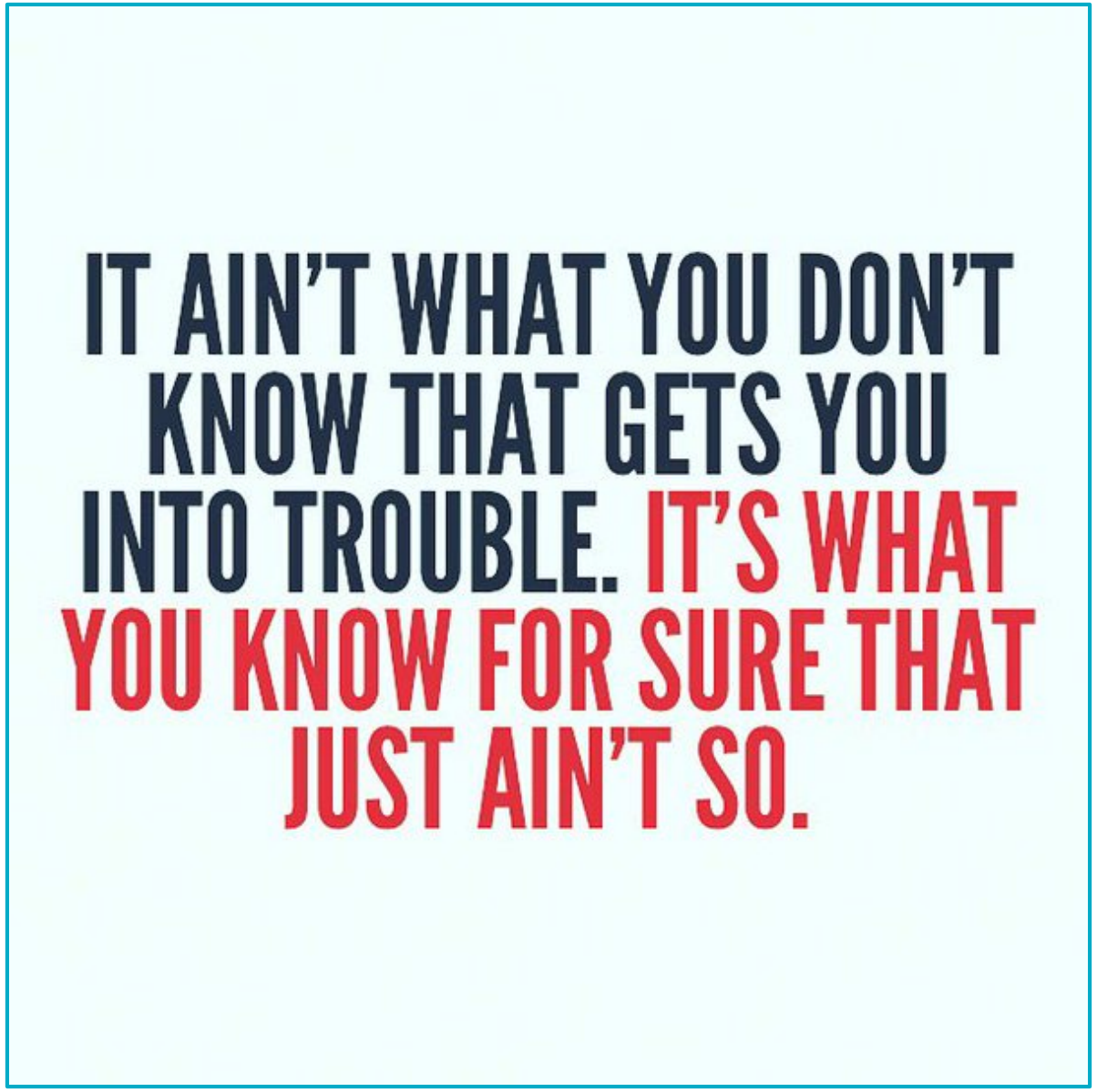
If you could sift through the guidelines, review papers, research articles and systematic reviews and then take the conclusions and publish a single synthesized answer…

YOU MAY discover new insights and YOU MAY provide the most useful information or…


YOU MAY amplify false signals
that are misinterpreted and
repeated across our medical
literature



--Stephen Hawking, quoting Daniel J. Boorstin

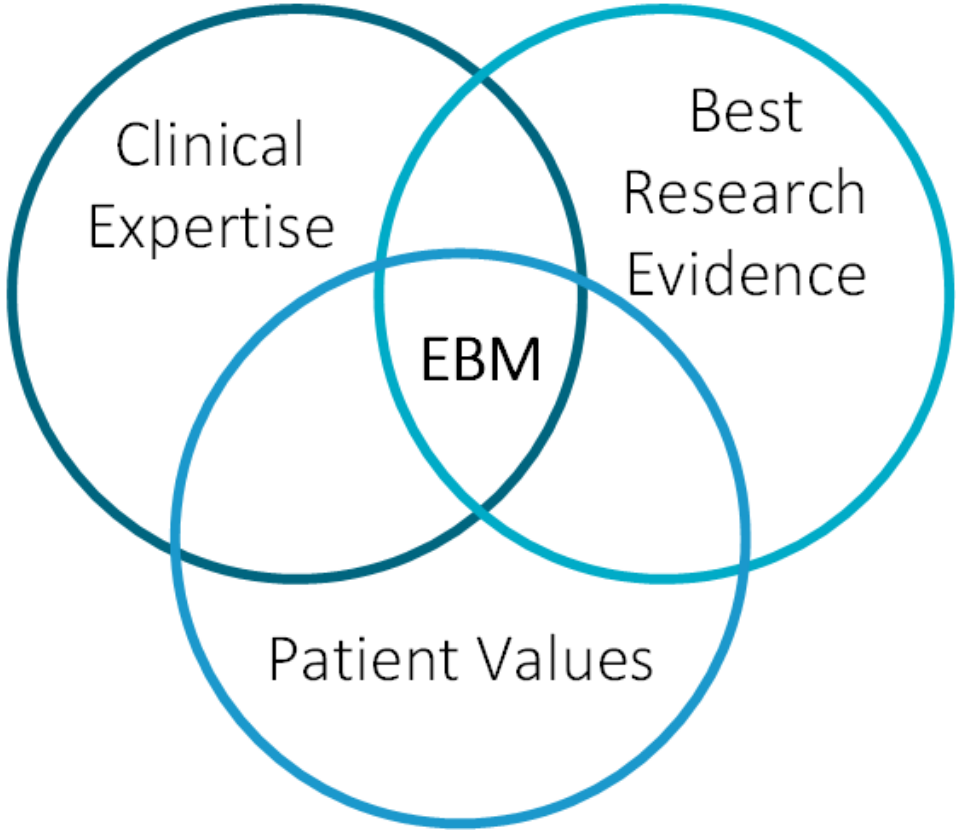


-- Mark Twain, quoting Josh Billings

Misunderstood “Best Research Evidence” can Cause Serious Harm

Once promoted “Life-saving treatments” that are now considered UNSAFE and INEFFECTIVE:

- Hormone replacement therapy for cardiovascular disease prevention
- Antiarrhythmics for premature ventricular complexes (PVCs) post-STEMI
- Perioperative beta blockers for STEMI prevention
- High-dose chemotherapy + bone marrow transplant for breast cancer



Best Research Evidence

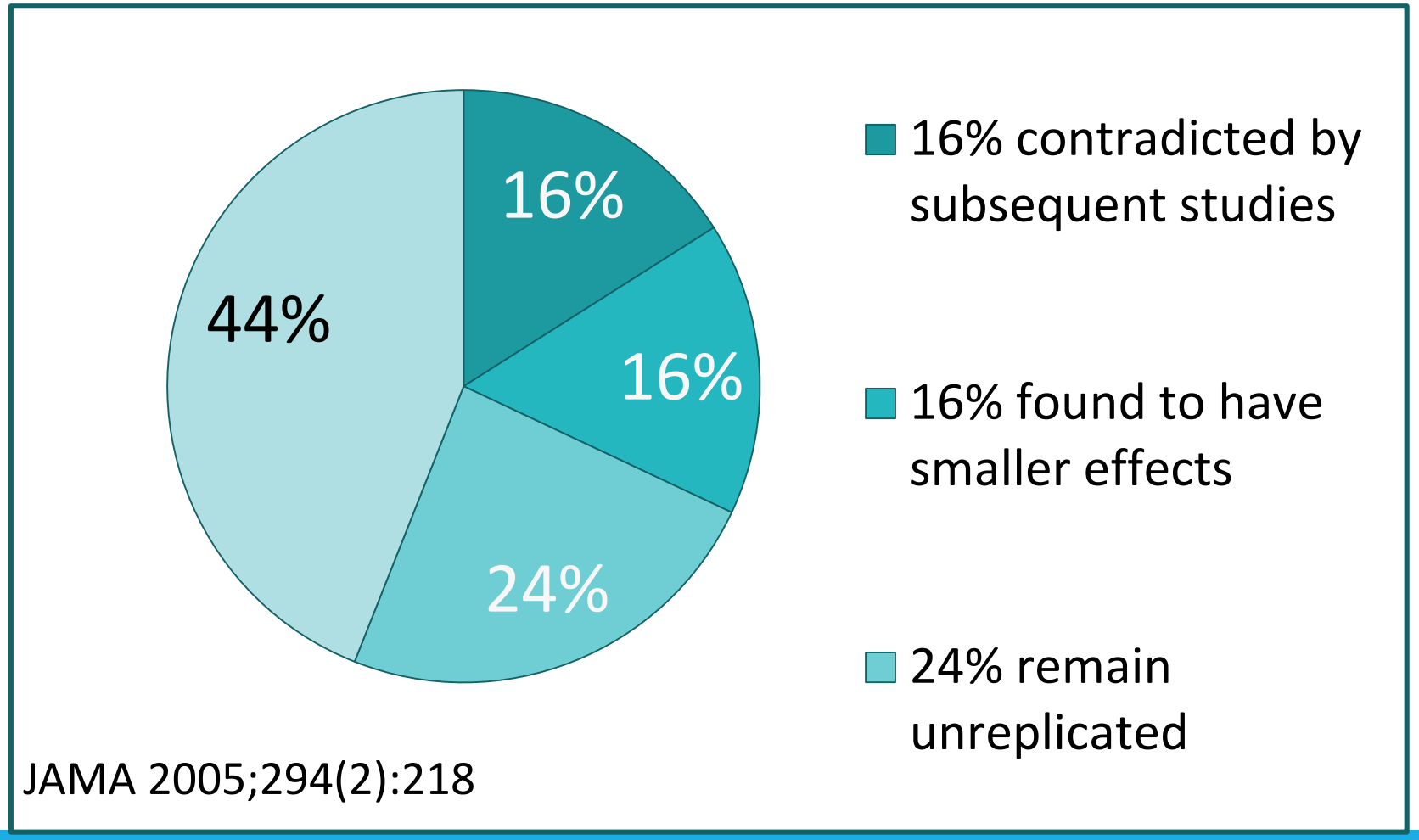
Comprehensive – Evidence can only be known to be best if all the available evidence known

Current – Every day research produces new evidence that could be best

Synthesized – Results from one study vs. the whole picture

Science requires replication of results for confirmation but …

Out of 45 highly-cited original research publications



JAMA 2005;294(2):218

Best Research Evidence

Comprehensive – Evidence can only be known to be best if all the available evidence known

Current – Every day research produces new evidence that could be best

Synthesized – Results from one study vs. the whole picture

Valid – Critical appraisal determines potential for bias

Why is critical appraisal essential?

What is published is often wrong, misleading, misinterpreted or incomplete

Selective summarization and citation from bias or familiarity

Interpretation of changes in surrogate markers to mean changes in clinical outcomes

Citation of what is published instead of tracing to original research

Acceptance and citation of conclusions of research instead of evaluating methods and statistics

Use of abstracts instead of full-text articles

Best Research Evidence

Comprehensive – Evidence can only be known to be best if all the available evidence known

Current – Every day research produces new evidence that could be best

Synthesized – Results from one study vs. the whole picture

Valid – Critical appraisal determines potential for bias

Systematic – Selection and evaluation of evidence by protocol reduces author bias, investigator bias, and editor bias

The EBM Solution: Systematic Review

Use systematic methods to achieve comprehensiveness, critical appraisal, and recognition and minimization of bias

Systematically…

…search for the evidence

…select the relevant evidence

…critically appraise the validity of the evidence

…combine evidence for interpretation of the body of evidence

But we do not have systematic reviews for most clinical questions, and most systematic reviews are outdated due to subsequently published evidence

Systematic Literature Surveillance
DynaMed 7-Step Methodology

	7-Steps to be Evidence Based	Systematic Review	Systematic Literature Surveillance
1 	Identifying the evidence	Systematic Search	Systematic Search
2 	Selecting the best available evidence	Does it answer the question?	**Does it answer a relevant question?**
3 	Critical Appraisal	Critical Appraisal	Critical Appraisal
4 	Objectively Reporting the Evidence	Objectively Reporting the Evidence	Objectively Reporting the Evidence
5 	Synthesizing Multiple Evidence Reports	Synthesizing Multiple Evidence Reports	Synthesizing Multiple Evidence Reports
6 	Basing Conclusions on the Evidence	Basing Conclusions on the Evidence	Basing Conclusions on the Evidence
7 	Maintaining Currency	Repeating search after ____ ?	**Updating Daily**

12 criteria for Level 1 [likely reliable] evidence for interventional conclusion (conclusions that an intervention does or does not change an outcome):

1. Full-text report available in English (or language well understood by participating editor)
2. Clinical outcome (also called patient-oriented outcomes)
3. Population, intervention, comparison, and outcome in the study is representative of expected clinical practice
4. Random allocation method (i.e. not assigned by date of birth, day of presentation, “every other”)
5. Blinding of all persons (patient, treating clinician, outcome assessor) if possible
6. Follow-up (endpoint assessment) of at least 80% of study entrants AND adequate such that losses to follow-up could not materially change the results
7. Accounting for dropouts (even if not included in analysis)
8. Confidence intervals do not include both presence and absence of clinically meaningful differences

12 criteria for Level 1 [likely reliable] evidence for interventional conclusion (conclusions that an intervention does or does not change an outcome):

9. In cases of randomized parallel-group trials
   i. Allocation concealment
   ii. Intention-to-treat analysis comparing groups according to randomization
10. In cases of randomized crossover trials
   i. 6 specific criteria (see website for details)
11. In cases of early trial termination
   i. 5 specific criteria (see website for details)
12. No other factors contributing to substantial bias, such as
   i. Differences in management between groups other than the intervention being studied
   ii. Differential loss to follow-up
   iii. Post hoc analysis
   iv. Subgroup analysis
   v. Baseline differences between groups
   vi. Unclear how missing data are accounted for

How Can a MACHINE Learn This?

Use the published evidence as is → GIGO → Amplify the problem

Limit “reference standard” for Machine Learning to “only the absolutely right” → inadequate Source Of Truth (most data is not absolute fact)

Limit “reference standard” for Machine Learning to “the most trustworthy sources” →

- Who decides what the most trustworthy sources are?
- How is the cutoff for trustworthiness determined?
- Do we still end up with inadequate Source Of Truth data to start?

Or will it take a large amount of human effort to build the reference standard to train the machine?

What Does the MACHINE Need to Learn?

We cannot answer “What are the effects (benefits and harms) of different management options?”

We can communicate precisely to the machine “What are the likely effects (benefits and harms) of different management options?”

But how we communicate our certainty or confidence in our likelihood or effect estimates is critical.

EBMonFHIR

A standard to communicate Evidence to the Machine

This is how the Machine will learn

Evidence Resource

...Metadata elements...			
referentGroup		1..1	BackboneElement
	description	0..1	markdown
	note	0..*	Annotation
	evidenceSource	1..1	Reference(Group | EvidenceSet)
	intendedGroup	0..1	Reference(Group)
	directnessMatch	0..1	CodeableConcept (Extensible: Low, Moderate, High, Exact)
variableDefinition		0..*	BackboneElement
	description	0..1	markdown
	note	0..*	Annotation
	variableRole	0..1	CodeableConcept (Extensible: exposure, measuredVariable)
	actualDefinition	1..1	Reference(EvidenceVariable)
	intendedDefinition	0..1	Reference(EvidenceVariable)
	directnessMatch	0..1	CodeableConcept (Extensible: Low, Moderate, High, Exact)
synthesisType		0..1	CodeableConcept
studyType		0..1	CodeableConcept
statistic		0..*	Statistic
distribution		0..*	OrderedDistribution
certainty		0..*	BackboneElement
	description	0..1	string
	note	0..1	Annotation
	rating	0..*	CodeableConcept
	certaintySubcomponent	0..*	BackboneElement
		description	0..1	string
		note	0..1	Annotation
		type	0..*	CodeableConcept
		rating	0..*	CodeableConcept