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.

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
THE GREATEST ENEMY OF KNOWLEDGE IS NOT IGNORANCE, IT IS THE ILLUSION OF KNOWLEDGE.

--Stephen Hawking, quoting Daniel J. Boorstin
IT AIN’T WHAT YOU DON’T KNOW THAT GETS YOU INTO TROUBLE. IT’S WHAT YOU KNOW FOR SURE THAT JUST AIN’T SO.

-- 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

- 16% contradicted by subsequent studies
- 16% found to have smaller effects
- 24% remain unreplicated

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

Use of abstracts instead of full-text articles

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

<table>
<thead>
<tr>
<th>7-Steps to be Evidence Based</th>
<th>Systematic Review</th>
<th>Systematic Literature Surveillance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identifying the evidence</td>
<td>Systematic Search</td>
<td>Systematic Search</td>
</tr>
<tr>
<td>2. Selecting the best available evidence</td>
<td>Does it answer the question?</td>
<td>Does it answer a relevant question?</td>
</tr>
<tr>
<td>3. Critical Appraisal</td>
<td>Critical Appraisal</td>
<td>Critical Appraisal</td>
</tr>
<tr>
<td>4. Objectively Reporting the Evidence</td>
<td>Objectively Reporting the Evidence</td>
<td>Objectively Reporting the Evidence</td>
</tr>
<tr>
<td>5. Synthesizing Multiple Evidence Reports</td>
<td>Synthesizing Multiple Evidence Reports</td>
<td>Synthesizing Multiple Evidence Reports</td>
</tr>
<tr>
<td>7. Maintaining Currency</td>
<td>Repeating search after ____?</td>
<td>Updating Daily</td>
</tr>
</tbody>
</table>
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 \(\rightarrow\) GIGO \(\rightarrow\) Amplify the problem

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

Limit “reference standard” for Machine Learning to “the most trustworthy sources” \(\rightarrow\)

  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