Validation and Regulatory Oversight of Clinical AI Tools

Session 7, March 10, 2020

Wade L. Schulz, MD, PhD
Assistant Professor, Yale School of Medicine

Harlan M. Krumholz, MD, SM
Professor, Yale School of Medicine

DISCLAIMER: The views, opinions and images expressed in this presentation are those of the author and do not necessarily represent official policy or position of HIMSS.
Meet Our Speakers

Wade Schulz
Assistant Professor, Yale School of Medicine

Harlan Krumholz
Professor, Yale School of Medicine
Wade Schulz, MD PhD

Salary: None
Royalty: None
Receipt of Intellectual Property Rights: Refactor Health from Yale University
Consulting Fees: HugoHealth, Interpace Diagnostics
Fees for Non-CME Services Received Directly from a Commercial Interest or their Agents (e.g., speakers’ bureau): None
Contracted Research: Shenzhen Center for Health Information
Ownership Interest (stocks, stock options or other ownership interest excluding diversified mutual funds): Refactor Health
Other: None
Conflict of Interest

Harlan Krumholz, MD SM

Salary: None
Royalty: None
Receipt of Intellectual Property Rights: Refactor Health from Yale University
Fees for Non-CME Services Received Directly from a Commercial Interest or their Agents (e.g., speakers’ bureau): None
Contracted Research: CMS, Medtronic, FDA, Johnson & Johnson, Shenzhen Center for Health Information
Ownership Interest (stocks, stock options or other ownership interest excluding diversified mutual funds): Refactor Health, HugoHealth
Other: Advisory Board for UnitedHealth, Element Science, Facebook, and Aetna; Participant IBM Watson Health Life Sciences Board
Agenda

• Importance RWD for the future of healthcare as an information science
• Emerging role of AI in healthcare analytics and decision support
• Importance of validation in clinical AI
• Regulatory oversight of AI and existing healthcare regulatory frameworks
• Best practices and the future of clinical AI validation, regulation, and implementation
Learning Objectives

• Describe current and future regulatory requirements for validation and ongoing quality assessment of predictive models used for clinical decision support
• Identify limitations of real-world data and approaches to reduce bias and error in machine learning models
• Explain best practices for implementation and local validation of artificial intelligence and machine learning models used for clinical decision support
Digital Revolution in Healthcare

• Healthcare finally experiencing digital revolution
Healthcare as an Information Science

• Data-driven approaches to diagnostic, therapeutic, and prognostic decision making

• Learning healthcare system where every interaction drives new knowledge that better informs choices and improves outcomes

• Performance metrics in real-time for accountability, improvement, choice
Healthcare Data Generation

- EHR
- Pathology / Laboratory
- Continuous Monitoring
- Payers / Claims
- Trials / Research
- Imaging
- Molecular Data
- Digital / Home Health
- Pharmacy
Digital Health Funding

TOTAL VENTURE FUNDING

<table>
<thead>
<tr>
<th>Year</th>
<th>Funding (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>$1.1B</td>
</tr>
<tr>
<td>2012</td>
<td>$1.5B</td>
</tr>
<tr>
<td>2013</td>
<td>$2.1B</td>
</tr>
<tr>
<td>2014</td>
<td>$4.1B</td>
</tr>
<tr>
<td>2015</td>
<td>$4.7B</td>
</tr>
<tr>
<td>2016</td>
<td>$4.6B</td>
</tr>
<tr>
<td>2017</td>
<td>$5.8B</td>
</tr>
<tr>
<td>2018</td>
<td>$8.2B</td>
</tr>
<tr>
<td>H1 2019</td>
<td>$8.4B</td>
</tr>
</tbody>
</table>

# OF DEALS

Average Deal Size:
- 2011: $12.0M
- 2012: $10.6M
- 2013: $10.7M
- 2014: $14.0M
- 2015: $14.7M
- 2016: $13.5M
- 2017: $15.9M
- 2018: $21.9M
- H1 2019: $23.1M

Evolution of Analytics

Descriptive

Diagnostic

Predictive

Prescriptive
Emerging Role for AI in Healthcare
Need for Guidance and Regulation

• Data are powerful, but can also lead to overload and misinterpretation.
Need for Oversight of AI

Becoming widely recognized as a need beyond healthcare
Regulatory Oversight of Clinical AI

Clinical Decision Support Software
Draft Guidance for Industry and Food and Drug Administration Staff

Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)

Discussion Paper and Request for Feedback
Food, Drug and Cosmetic Act

“The FDA has long regulated software that meets the definition of a device in section 201(h) of the Federal Food, Drug and Cosmetic Act”

Includes: “software that is intended to provide decision support for the diagnosis, treatment, prevention, cure or mitigation of diseases or other conditions (often referred to as clinical decision support [CDS] software)”
What is classified as decision support?

- **Computerized alerts and reminders for providers and patients**
  - Clinical guidelines
  - Condition-specific order sets

- **Focused results and patient data**
  - Highlighted patient reports and summaries
  - Documentation templates

- **Diagnostic support**
  - Contextually-relevant reference information
  - Clinical diagnostic support tools
When is decision support a medical device?

- Is NOT “intended to acquire, process, or analyze a medical image or a signal from an in vitro diagnostic device or a pattern or signal from a signal acquisition system”
- Is “intended for the purposes of displaying, analyzing or printing medical information about a patient or other medical information (such as peer-reviewed clinical studies and clinical practice guidelines)”
When is decision support a medical device?

- Is “intended for the purpose of supporting or providing recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition:

- AND NOT “intended for the purposes of enabling such health care professional to independently review the basis for such recommendations such that software presents so that it is not the intent that such health care professional rely primarily on any such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient”
# CDS as a Medical Device

## Table 1. Is a CDS Software Function Device or Non-Device?

<table>
<thead>
<tr>
<th>Is the Intended User an HCP? [part of criteria (3) and (4)]</th>
<th>Can the User Independently Review the Basis?* [part of criterion (4)]</th>
<th>Is it Device CDS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No, it is Non-Device CDS because it meets all of section 520(a)(1)(E) criteria</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes, it is Device CDS</td>
</tr>
<tr>
<td>No, it is a patient or caregiver</td>
<td>Yes</td>
<td>Yes, it is Device CDS</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes, it is Device CDS</td>
</tr>
</tbody>
</table>

*“Can the user independently review the basis?” asks whether the function is intended for the purpose of enabling the user to independently review the basis for the recommendations so that it is not the intent that user rely primarily on any such recommendation (part of criterion (4)).
## Software as a Medical Device

![Image: Figure 1: SaMD IMDRF risk categorization](image)

<table>
<thead>
<tr>
<th>State of healthcare situation or condition</th>
<th>Significance of information provided by SaMD to healthcare decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treat or diagnose</td>
</tr>
<tr>
<td>Critical</td>
<td>IV</td>
</tr>
<tr>
<td>Serious</td>
<td>III</td>
</tr>
<tr>
<td>Non-serious</td>
<td>II</td>
</tr>
</tbody>
</table>

*Figure 1: SaMD IMDRF risk categorization*
# FDA Approved Algorithms

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Number of FDA Approved Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>1</td>
</tr>
<tr>
<td>2015</td>
<td>0</td>
</tr>
<tr>
<td>2016</td>
<td>4</td>
</tr>
<tr>
<td>2017</td>
<td>8</td>
</tr>
<tr>
<td>2018</td>
<td>25</td>
</tr>
<tr>
<td>2019 (as of June)</td>
<td>8</td>
</tr>
<tr>
<td>Total (as of June 2019)</td>
<td>46</td>
</tr>
</tbody>
</table>

Source: The Medical Futurist (6 June 2019) FDA Approvals For Smart Algorithms In Medicine In One Giant Infographic.
Software as a Medical Device
What to Validate in a Clinical AI Model

- Data Quality
- Clinical Phenotype / Outcome
- Analytic and Clinical Efficacy
- Ongoing Accuracy

Risk of Bias
Secondary Use of EHR Data

1. **Are data fit-for-purpose?**
   Main role of EHR is clinical documentation and direct provision of care, despite evolution of EHR leading to increased scope and use.

2. **Can data be generalized?**
   Even within a single system, may be a multitude of data standards, laboratory methodologies, and EHR customizations.

3. **Are data available when needed?**
   Data such as unstructured clinical notes may be available for retrospective analysis, but not immediately available for clinical predictive models.
AI as an in silico diagnostic

Data Quality

Clinical Phenotype / Outcome

Analytic and Clinical Efficacy

Ongoing Accuracy

Pre-Analytical

Analytical

Post-Analytical

#HIMSS20
Data Quality

• Issues with data ‘quality’ can be diverse and arise at many stages
  • Missing data or data in wrong location
  • Changes to coding practices
  • Implementation of new laboratory methodology
  • Variation in provider workflow
Data Quality – Features and Labels
Phenotype / Cohort Identification

![Graph showing the probability of missing diagnosis over years from second signal for HTN, HLD, and DM cases.](https://www.medrxiv.org/content/10.1101/2019.12.28.19015628v1)
AI as an in silico diagnostic

- Data Quality
- Clinical Phenotype / Outcome
- Analytic and Clinical Efficacy
- Ongoing Accuracy
- Pre-Analytical
- Analytical
- Post-Analytical
# In Vitro Diagnostics: Test Validity

<table>
<thead>
<tr>
<th>Analytical</th>
<th>Clinical Validity</th>
<th>Clinical Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>How well a test performs at detecting a given analyte. Analytically valid tests are precise, accurate, and reliable.</td>
<td>How accurately a test predicts the presence of or risk for a specific condition.</td>
<td>Whether a test is associated with improved outcomes or more efficient treatment.</td>
</tr>
</tbody>
</table>
Analytical Efficacy – Beyond the C-Statistic
Correlation to Laboratory Test Validation

1. Accuracy (Method Correlation)
2. Clinical Sensitivity (Positive Predictive Value)
3. Clinical Specificity (Negative Predictive Value)
4. C-Stat
5. F1 Score
6. Precision (Repeatability/Reproducibility)
7. Reportable Range (Linearity)
8. Analytical Sensitivity (LOD;LOQ)
9. Analytical Specificity (Interferences)
10. Reference Range (Reference Interval)
FDA vs CMS

• Along with the specific regulation of medical devices, oversight of clinical diagnostic testing can fall under either FDA or CMS
  • CMS statutory authority under Clinical Laboratory Improvement Amendments (‘88); FDA under Medical Device Amendments (‘76)
  • All in vitro diagnostics are medical devices and fall to FDA, but has not historically exercised regulatory authority over laboratory developed tests (LDTs)
  • If tests not subject to FDA clearance or approval, laboratory must establish performance characteristics of the test
Clinical Utility – A Case Study

• Goal to create a predictive model to predict bed utilization

• Clinical unit that is typically at / near capacity

• Model developed with excellent performance characteristics, which continues to predict unit to be at / near capacity
  • Limited clinical utility, even with an accurate prediction, limited ability to implement meaningful changes to clinical workflows
AI as an in silico diagnostic

- Data Quality
- Clinical Phenotype / Outcome
- Analytic and Clinical Efficacy
- Ongoing Accuracy
- Pre-Analytical
- Analytical
- Post-Analytical
Good Machine Learning Practices & Ongoing Validation

Analytical model management
Model input overview
Model change/version control

Staff Training & Certification
Competency assessments
Documentation & standards
Privacy and bias training

Data and result validation
Quality control assessment
Proficiency testing assessment
Clinically-Valid Analytics and AI Pipelines: SaMD
Conclusions: The Future of Healthcare AI

- Opportunity to revolutionize the mechanism by which evidence is generated and healthcare is delivered
- Need for iteration and development in a rapidly evolving field, but critical to take a measured approach in clinical implementation
- Regulation can stifle innovation, but should be seen as a mechanism to guarantee safety, limit bias, and improve efficacy of clinical analytics
- Existing frameworks can be used to guide clinical AI development and implementation
Questions

- Wade Schulz (wade.schulz@yale.edu @wade_schulz)
- Harlan Krumholz (harlan.krumholz@yale.edu @hmkyale)

- This slide provides a link that will take you directly to the evaluation form:
  <link>
RWD: A New Era for Evidence Generation

- Current methods of evidence generation are too slow
- To support precision medicine and advanced decision support, need large cohorts to identify specific phenotypes
- Comprehensive, digital health records provide opportunity to drive innovation and identify outcomes that could not be obtained before
Real-World Data

Randomized Trial
- Controlled setting
- Placebo / control intervention
- Homogeneous population
- Standardized follow-up

Real-World Data
- Generalized setting
- Clinical practice
- Variable population
- Variable, provider-dependent follow-up