ENHANCING FDA REGULATION OF AI/ML ALGORITHMS

David H. Jiang, B.A.

CERSI Scholars Virtual Seminar
Friday, May 7, 2021
ACKNOWLEDGEMENTS

Yale-Mayo CERSI
James Han
Lindsay Emanuel
Sanjay Aneja, M.D.
Christopher Morten, J.D., Ph.D.
Dennis Shung, M.D., M.H.S.
Peter Noseworthy, M.D.
Joseph Ross, M.D., M.H.S.
Nilay Shah, Ph.D.
## BACKGROUND
### FDA RISK AND APPROVAL PROCESS

### FDA Device Classification

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>Devices with minimal risk; are generally exempt from regulatory review beyond general controls</td>
</tr>
<tr>
<td>Class II</td>
<td>Devices for which general controls are insufficient to provide reasonable assurance of the safety and effectiveness</td>
</tr>
<tr>
<td>Class III</td>
<td>Devices are those devices that usually sustain or support life, implanted, or present a potentially unreasonable risk of injury to the patient; requires safety &amp; effectiveness to be demonstrated</td>
</tr>
</tbody>
</table>

### FDA Device Approval

<table>
<thead>
<tr>
<th>Approval Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>510(k) Clearance</td>
<td>Manufacturer must prove “substantial equivalence” to a previously approved device</td>
</tr>
<tr>
<td>Premarket Approval</td>
<td>Used for Class III high risk devices. Requires rigorous proof that the device is safe and effective</td>
</tr>
<tr>
<td>De Novo Pathway</td>
<td>New devices without a predicate; FDA decides of risk levels.</td>
</tr>
</tbody>
</table>
BACKGROUND
PREVIOUS STUDIES

There has been two studies that made preliminary steps into this research.

• Benjamens et al.
  • Looked at 29 AI devices identified through searching summaries
  • Found an additional 35 through press releases

• Wu et al.
  • Looked at 130 AI devices
  • Assigned risk scores
  • Contained some study characteristics


STUDY AIMS

AIM 1

To build a more complete database of FDA approved AI/ML devices.

AIM 2

Identify and evaluate the measures of testing, trials, and data used to approve AI/ML SaMDs through the 510k, PMA, and De Novo processes.
METHODS

DATASET CONSTRUCTION

- 510(k) Clearance
- Premarket Approval
- De Novo Pathway

Check for Keyword Indicators

- "artificial intelligence"
- "machine learning"
- "neural networks"
- "deep learning"
- "reinforcement learning"
- "natural language processing"
- "decision tree"
- "Bayesian network"

Manual/Crosscheck

Resultant database
METHODS
DATA ABSTRACTION

• We read through each individual FDA summary and abstracted the following information, if available;
  • Country of Origin
  • Method of Approval
  • Study size
  • Non-clinical testing
    • Performance testing (i.e., validation & verification, bench testing, etc.)
    • Risk analysis (i.e., risk management)
    • Standards Utilization (i.e., industry standards, voluntary standards, etc.)
  • Clinical testing
    • Sensitivity & Specificity
    • Time to notification/time to open
    • Human factors/quality testing
RESULTS

510(k) Clearance
83,083

Premarket Approval
1,568

De Novo Pathway
303

Since 01/01/2010

510(k) Clearance
34,148

Premarket Approval
236

De Novo Pathway
251

Check for Keyword Indicators

208

Manual/Crosscheck

145

Resultant database

182

Benjamens et al.
Wu et al.
37
## RESULTS
### SUMMARY

- 135 Unique Devices
- 104 Unique Manufacturers

<table>
<thead>
<tr>
<th>Approval Process</th>
<th>Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>510 (k) Clearance</td>
<td>173</td>
</tr>
<tr>
<td>De Novo</td>
<td>9</td>
</tr>
<tr>
<td>Premarket Approval</td>
<td>0*</td>
</tr>
</tbody>
</table>

*Note: Premarket Approval value is marked with an asterisk.*
RESULTS

SUMMARY

- Radiology: 78%
- Cardiovascular: 11%
- Neurology: 4%
- Hematology: 3%
- Ophthalmologic: 2%
- Other: 2%
- Devices: N=182
RESULTS

DEVICE TESTING

Identified Devices
182

Clinical
88

Non-Clinical
94

Performance Testing
156

Risk Analysis
83

Standards Utilization
95

Change Protocol
1
RESULTS
CLINICAL TESTING

Study Design

Clinical Testing 88
Retrospective 81
Prospective 7

Characteristics

Clinical Trial Reg. 6
Multisite 39
Median = 3.5
Study Size 83
Median = 286

Evaluation Parameters

Sensitivity & Specificity 59
Time to notification 16
Human factor 9
DISCUSSION

• Vast majority of the devices are approved through 510(k) Notifications
  • Virtually none with PMA
• Radiology is the largest medical specialty
• Only 48% of devices (88) had some form of testing with data
  • Out of which the majority 92% (81) were retrospective testing
• No change protocol
• It is not specifically known the nature of the AI program in each device.
• No real outcomes tested/evaluated
DISCUSSION
POLICY IMPLICATIONS

AI Identification
AI Devices should be identified as such

Prospective Study
AI’s approval process should involve prospective clinical studies

Clinically Relevant Outcomes
Endpoints and outcomes should be clinically significant and relevant, not just a comparison between devices

Change Protocol
AI Devices should clearly define the planned course of change, especially for “unlocked” AIs.
DISCUSSION

FUTURE RESEARCH
CONCLUSION

By constructing a database of FDA approved AI/ML SaMDs, we were able to give a preliminary look into the measures of testing, trials, and data used to approve these devices. We recommended specific policy changes to better ensure transparency, safety and effectiveness of AI/ML devices and provided directions for future research.
QUESTIONS?
JIANG.DAVID@MAYO.EDU