Regulation and Assessment of Quantitative Imaging Tools Used in Precision Oncology

Nicholas Petrick. Ph.D.

Division of Imaging, Diagnostics and Software Reliability (DIDSR)
Office of Science and Engineering Labs
Center for Devices and Radiological Health (CDRH)
U.S. Food and Drug Administration (FDA)
nicholas.petrick@fda.hhs.gov

DISCLOSURES

- None

OUTLINE

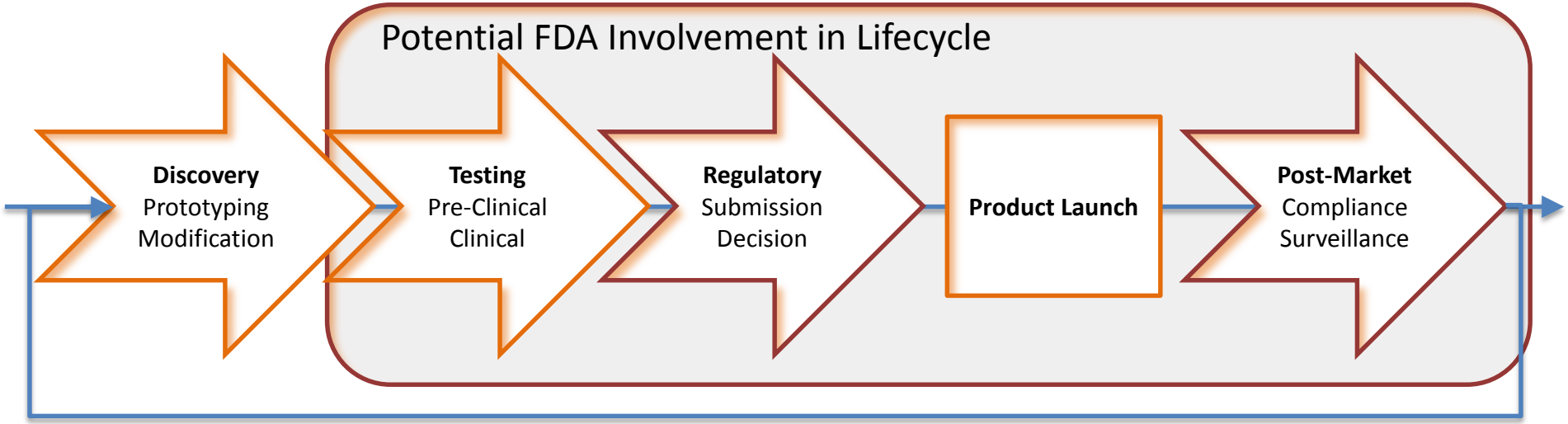
- Overview of medical device regulatory framework

- Quantitative imaging tool assessment
  - Lung nodule volumetry

Center for Devices and Radiological Health

- Protect and promote the health of the public by ensuring the **safety** and **effectiveness** of medical devices and the safety of radiation-emitting electronic products

Development Pathway



- FDA strives to speed translation of innovative, safe, and effective products to market throughout product lifecycle

Device Class & Pre-Market Requirements

Device Class	Controls	Premarket Review Process
Class I (lowest risk)	General Controls	Most are exempt
Class II	General Controls Special Controls	Premarket Notification [510(k)] or De Novo
Class III (highest risk)	General Controls Premarket Approval	Premarket Approval [PMA]

General/Special Controls

- General Controls
  - General controls apply to all medical devices, unless exempted by regulations
    - Registration and device listing
    - Adverse event reporting
    - Good manufacturing practice requirements
    - …
- Special Controls
  - Controls beyond general controls necessary to establish a reasonable assurance of the safety & effectiveness. Special controls are usually device-specific
    - Postmarket surveillance
    - Special labeling requirements
    - Premarket data requirements
    - …

http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/GeneralandSpecialControls/ucm055910.htm

How devices come to market in U.S.

- 510(k)
  - Demonstrate substantial equivalence to predicate device
- De Novo
  - Risk-based classification for novel medical devices for which general controls, or general and special controls, provide reasonable assurance of safety and effectiveness for the intended use, but for which there is no legally marketed predicate. Devices granted through De Novo may be marketed/used as predicates for future 510(k) submissions
- PMA
  - Demonstrate reasonable assurance of safety and effectiveness
  - Most Class III devices
- Qsubs
  - Informal interaction with FDA (usually non-binding) prior to device submission
    - Answer questions about a specific device under development

Medical Devices By Class










Class I

Class II

CT, MR, US imaging systems
Most imaging CADe/CADx
Some IVD tests

Class III

Novel Imaging systems (DBT)
Leadless Pacemakers
Bronchial Thermoplasty Systems
Some IVD Tests

REGULATION OF SAMD

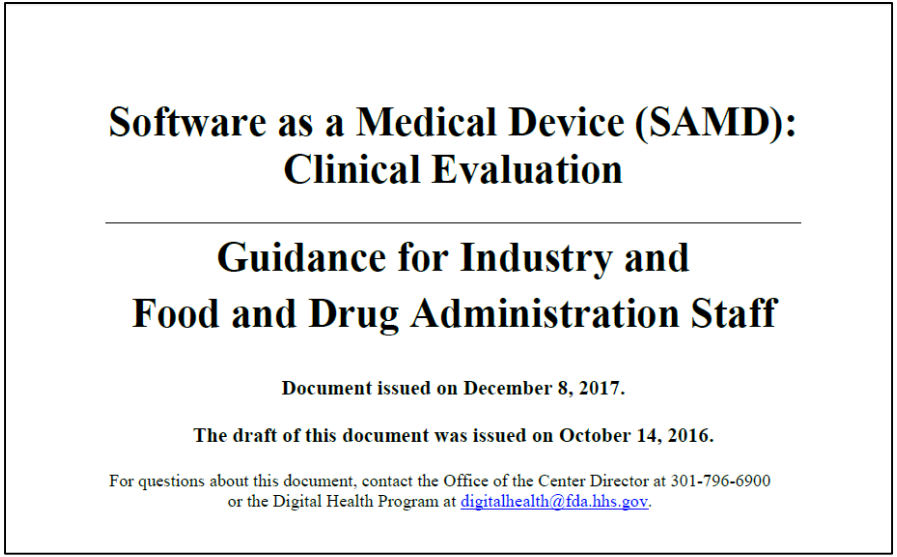
- IMDRF Working Group (WG) on Software as a Medical Device (SaMD)
  - SaMD: Software intended to be used for medical purposes without being part of a hardware medical device
    - Include artificial intelligence (AI) algorithms for disease diagnosis & monitoring
      - Including precision oncology tools
  - Outputs:
    - SaMD: Key Definitions
    - SaMD: Possible Framework for Risk Categorization and Corresponding Considerations
    - SaMD: Application of Quality Management System
    - SaMD: Clinical Evaluation

http://www.imdrf.org/workitems/wi-samd.asp

IMDRF International Medical Device Regulators Forum

IMDRF AND FDA GUIDANCE

- SAMD: Clinical Evaluation
  - Adopted as FDA guidance in 2017
  - FDA intends to consider principles of the IMDRF report in evolving approach to AI/ML and SaMD review



Software as a Medical Device (SAMD): Clinical Evaluation

Guidance for Industry and Food and Drug Administration Staff

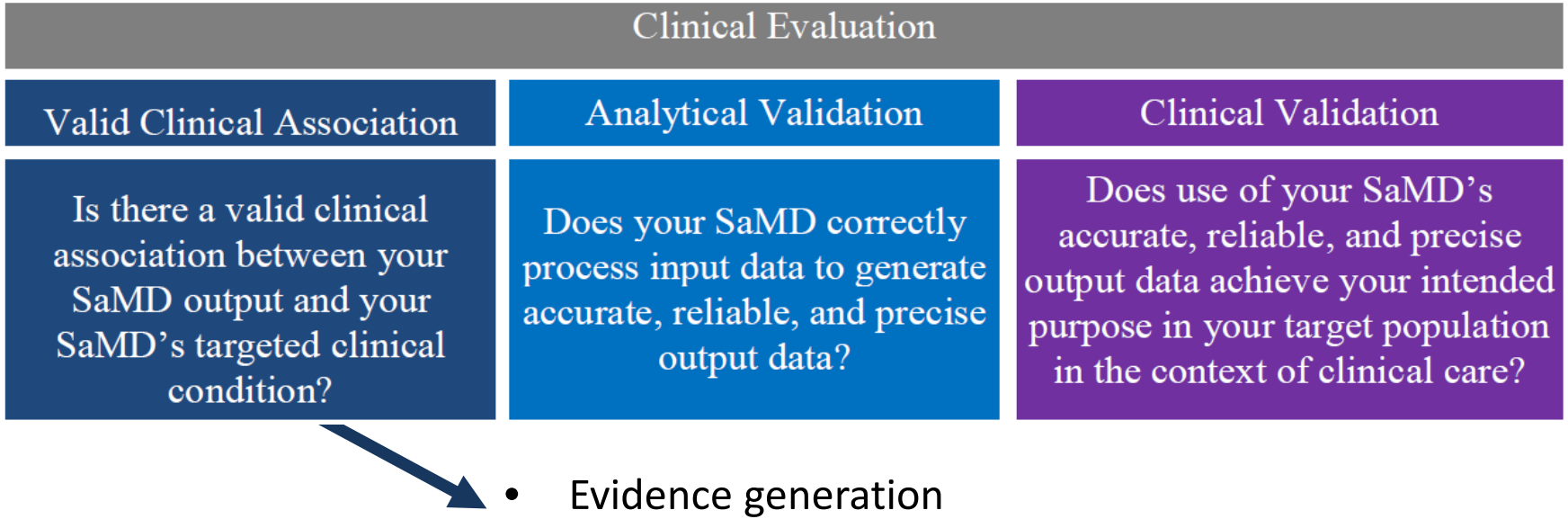
Document issued on December 8, 2017.

The draft of this document was issued on October 14, 2016.

For questions about this document, contact the Office of the Center Director at 301-796-6900 or the Digital Health Program at digitalhealth@fda.hhs.gov.

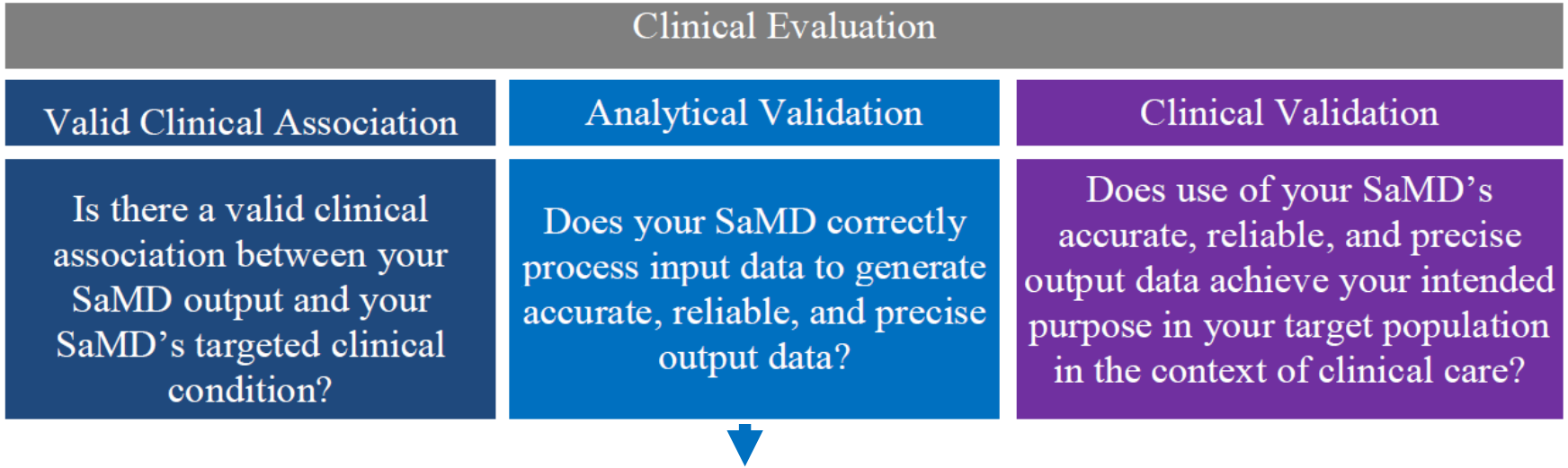
https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm524904.pdf

SaMD: Clinical Evaluation



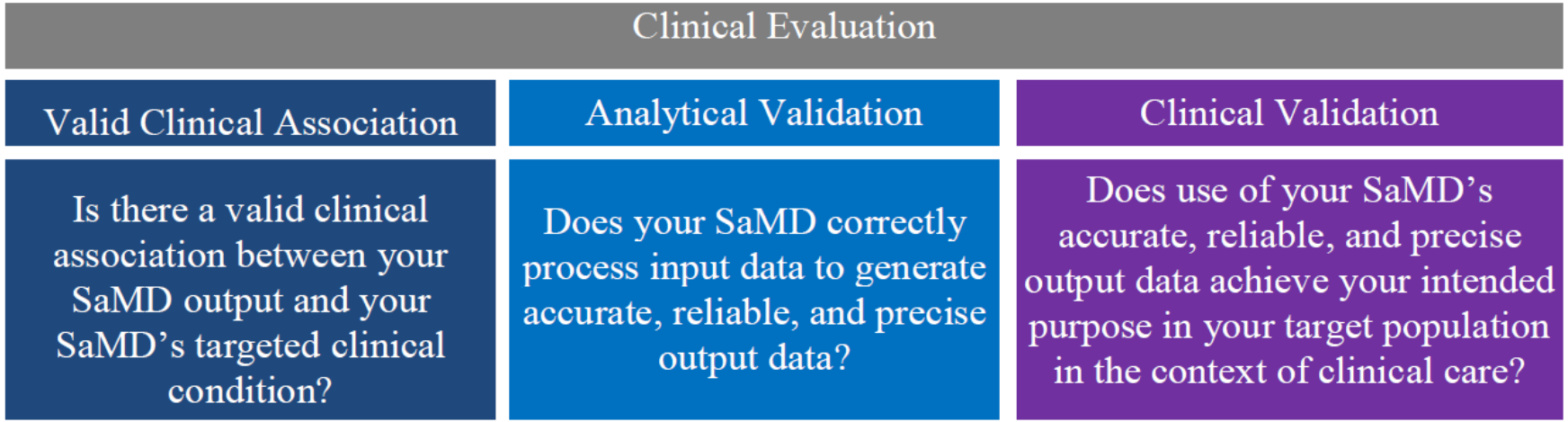
- Evidence generation
  - Literature
  - Professional guidelines
  - Secondary data analysis
  - Clinical trials/studies

SaMD: Clinical Evaluation



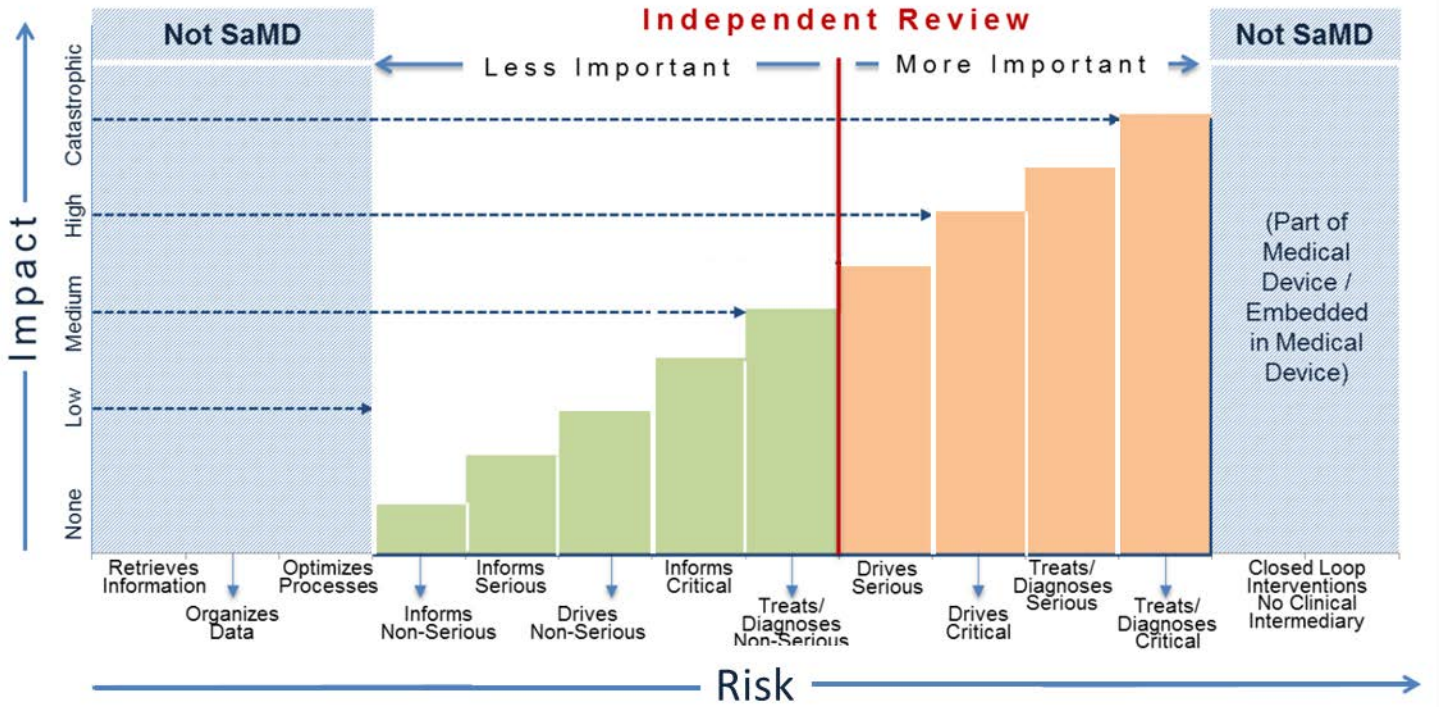
- SaMD meet technical requirements
  - Provide evidence that software correctly constructed
  - Demonstrate it meets specifications and conforms to user needs

SaMD: Clinical Evaluation



- Evidence that shows
  - SaMD has been tested in target population and for intended use
  - Users can achieve clinically meaningful outcomes

Risk-Based Regulatory Approach



Device Regulation Takeaways

- Devices are classified into three tiers
- Indications for use and type of technology are equally important for deciding what validation is needed
- You can ask FDA questions in a pre-submission (Qsub)
- AI tools
  - FDA has substantial guidance on AI tool assessment
  - FDA's approach to AI/SaMD is now evolving

QUANTITATIVE IMAGING TOOL ASSESSMENT

Quantitative Imaging

- Extraction of quantifiable features from medical images for the assessment of normal or the severity, degree of change, or status of a disease, injury, or chronic condition relative to normal
  - Single feature
  - Multiple features (artificial intelligence)
- Center working on QI guidance document

EXAMPLE CASE: CT VOLUMETRY

- CT lesion volume is a quantitative measure of actual tumor size in vivo
  - Actual quantitative imaging biomarker (QIB) claims would add specific performance goals to be achieved
  - Performance across multiple QI tools

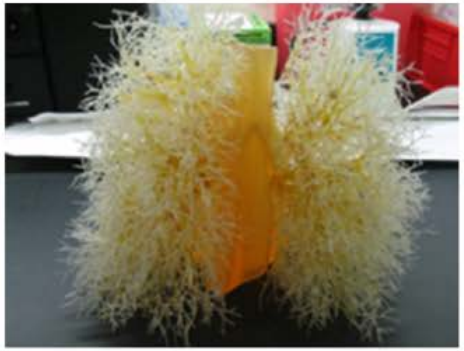
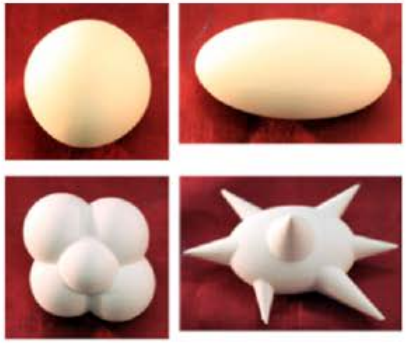
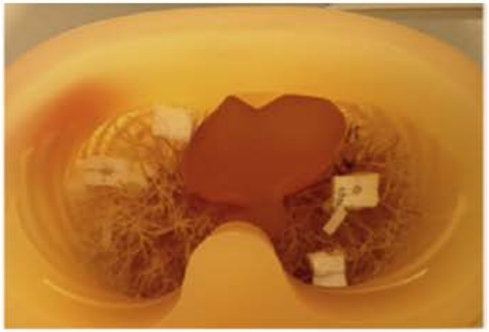
QI Tool Assessment

- Basic components
  - Verification and Validation (V&V) software testing
    - Does tool meet technical specifications
    - Analytical validation
      - Technical assessment
  - Clinical assessment
    - Typical requires specific randomized clinical studies
      - Generally need patient outcome data

Technical Assessment

- Statistical measures of performance
  - Accuracy
    - Linearity
    - Bias
  - Precision
    - Repeatability
    - Reproducibility

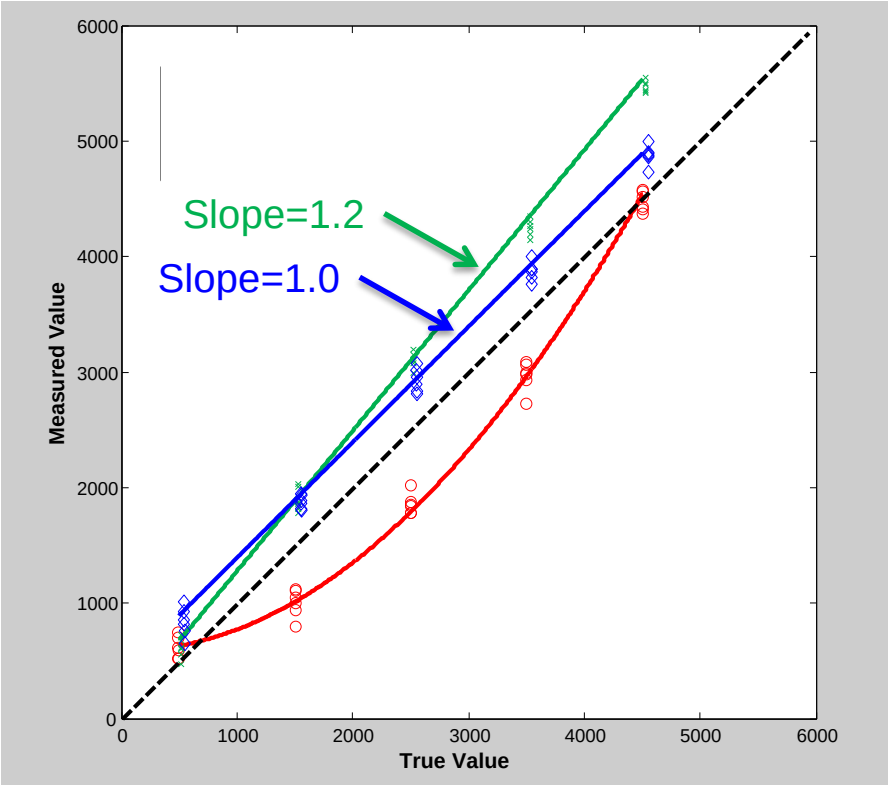
TECHNICAL ASSESSMENT-ACCURACY

- Accuracy assessment requires truth
- No volume reference from clinical CT scans
  - Phantom-based study
    - Thorax phantom with synthetic nodules

LINEARITY

- Measure of how change in reference reflects proportional change in measurement on avg.
  - Linear:
    Proportional change (straight line)
    - Blue: Slope 1.0
    - Green: Slope 1.2
  - Nonlinear
    - Red: Non-proportional change



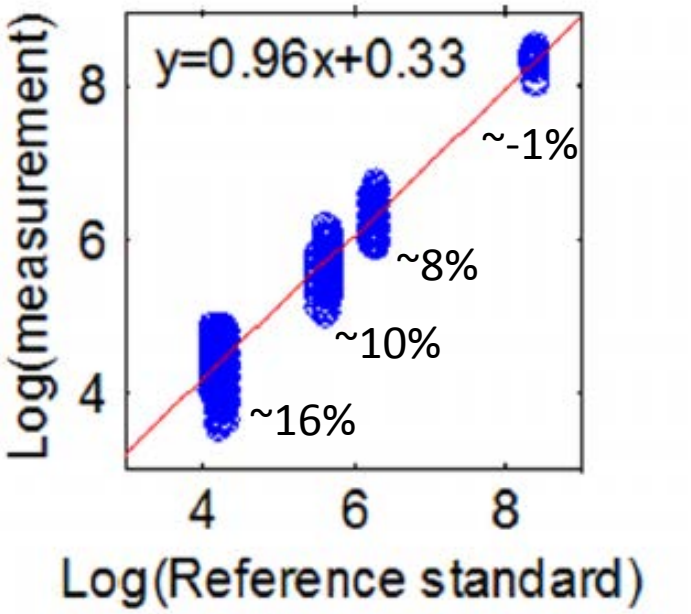
BIAS

- Difference between an estimator's expected value & reference standard
  - Scaler value

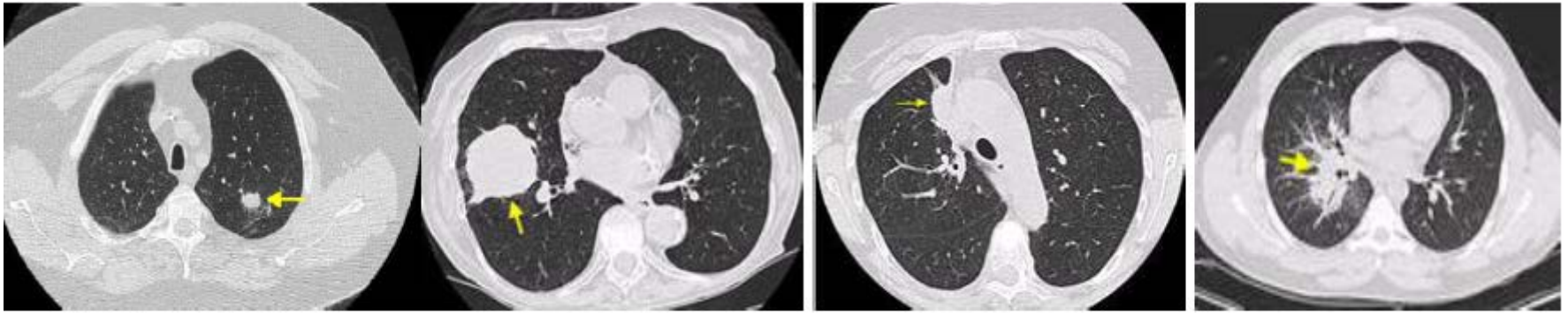
- Types
  - Unbiased
    - Bias=0
  - Constant bias
    - Bias=B
  - Non-constant bias
    - Bias=function of true value

PHANTOM STUDY LINEARITY ANALYSIS

- Accuracy assessment
  - Linearity
    - y= 0.96x+0.33
      - R^2= 0.98
      - Slope= 0.96 [0.96, 0.96]
  - Bias
    - B= 39% [38.1,40.1]



TECHNICAL ASSESSMENT-PRECISION



- Repeatability/reproducibility assessment
  - Clinical coffee-break scans
    - Patient scanned twice in <15 minutes
      - Same scanner
      - Same reconstruction

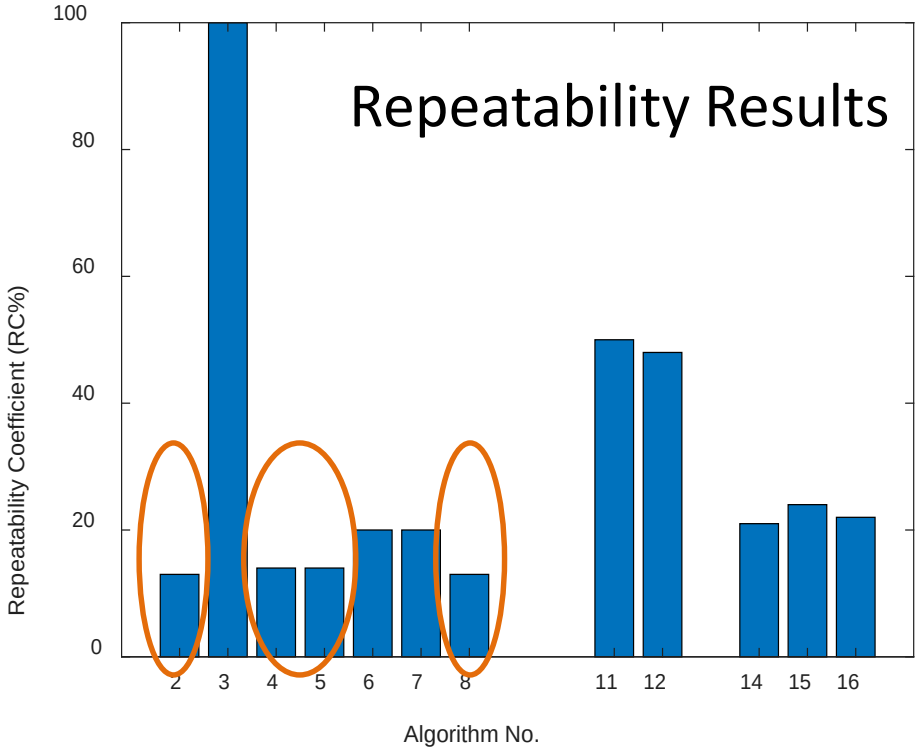
*Buckler et al., Acad Radiol, 2015.

TECHNICAL ASSESSMENT-PRECISION

- Clinical repeatability
  - Variability under identical or near-identical conditions
    - Individual algorithm assessment
    - Repeatability coefficient (RC)
      - smaller RC → higher precision
- Clinical reproducibility
  - Variability under a range of experimental conditions
    - Across multiple algorithm
    - Reproducibility coefficient (RDC)
      - smaller RDC → higher precision

Clinical Repeatability Analysis

- Repeatability
  - Single algorithm
  - RC range: 13%-100%
  - Best performers
    - ~13-14%
      - Algorithms: 2,4,5,8
    - Interchangeable?



*Buckler et al., Acad Radiol, 2015.

CLINICAL REPRODUCIBILITY ANALYSIS

Algorithm Grouping	RDC	RDC%
All but Algm 3	25,284 mm^3	84%
Best Performers	9,290 mm^3	58%

- Reproducibility
  - Across algorithms
  - Even best algorithms pay a large penalty if interchanged
    - 13% → 58%

*Buckler et al., Acad Radiol, 2015.

SUMMARY

- Device Regulation
  - Devices are classified into three tiers
  - Indications for use and type of technology are equally important for deciding what validation is needed
  - FDA's approach to AI/SaMD is now evolving

SUMMARY

- QI/radiomic tool assessment
  - Basic components
    - Verification and Validation (V&V) testing
      - Technical assessment
    - Clinical assessment

ACKNOWLEDGMENTS

- I’ d like to acknowledge Qin Li, Marios Gavrielides, and Ben Berman for their QI research efforts presented in this talk


FDA
U.S. FOOD & DRUG
ADMINISTRATION