Protocol Development and Quality Assurance Steps to Ensure Data Integrity

Susanna I Lee MD, PhD

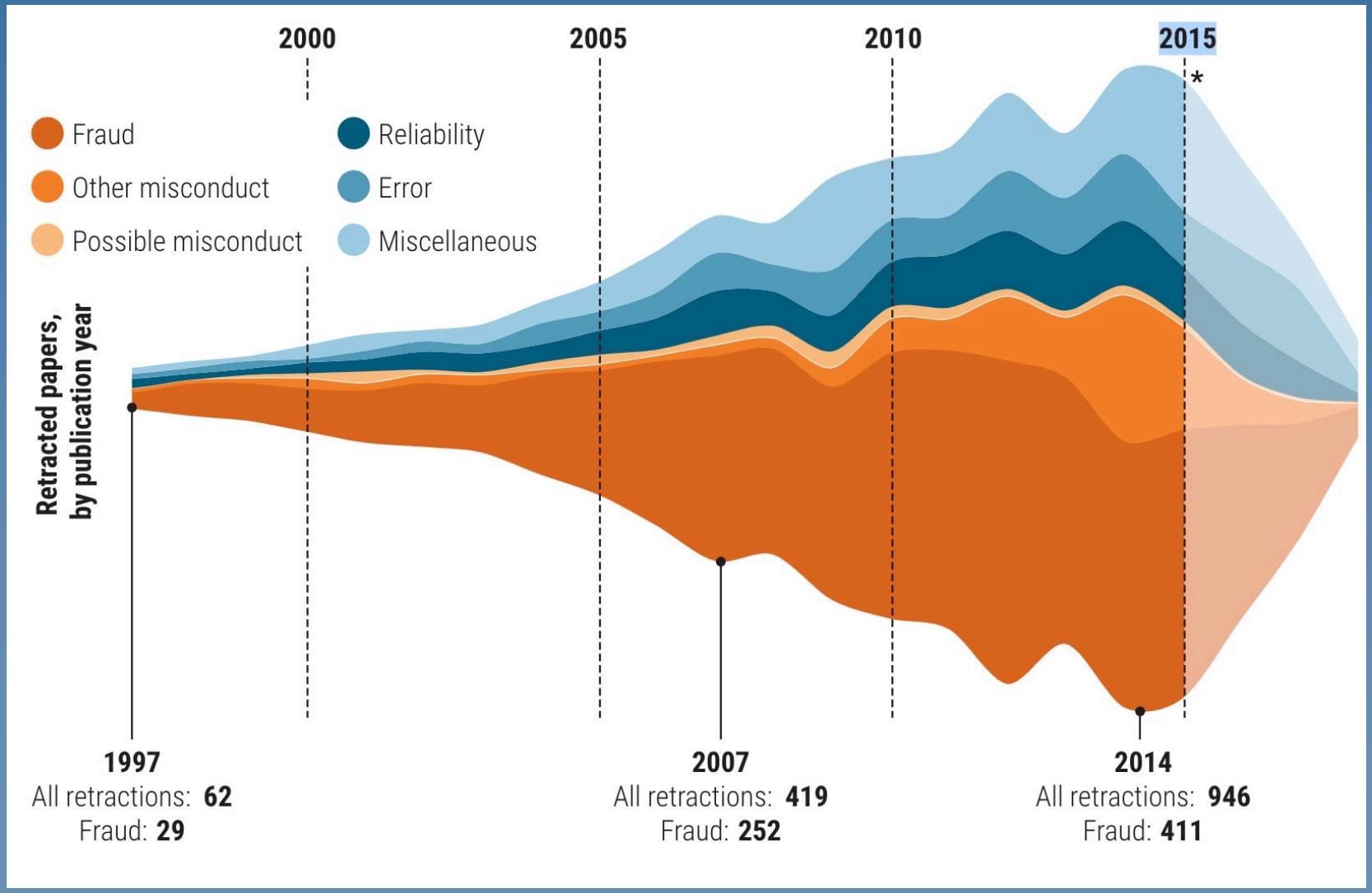
Disclosures

- **Relevant: None**
- **Other: Royalties Wolters Kluwer, Springer, RSNA**

Learning Objective

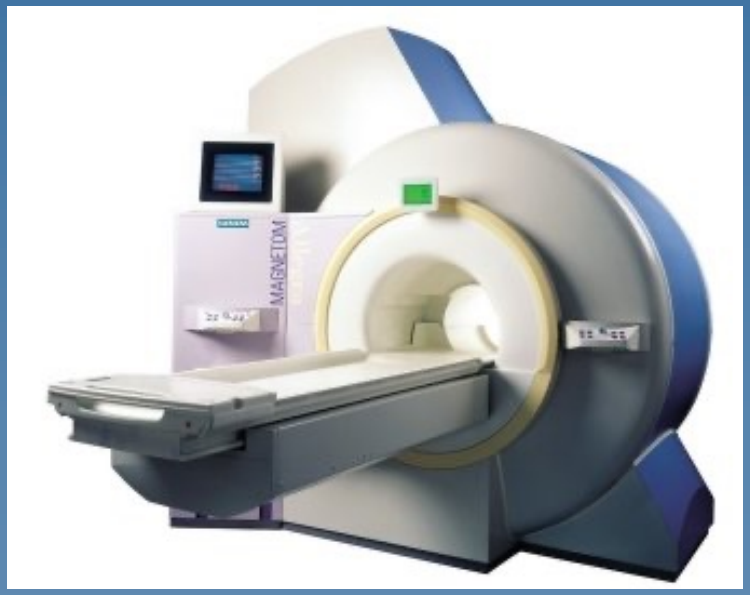
- Anticipate and manage sources of data imprecision
- Collect databases structured for ongoing QA, interim and final analysis, and retrieval for future use

Retracted Scientific Publications



https://www.science.org/content/article/what-massive-database-retracted-papers-reveals-about-science-publishing-s-death-penalty

Imaging Chain: Patient → Data



Raw data

Image reconstruction

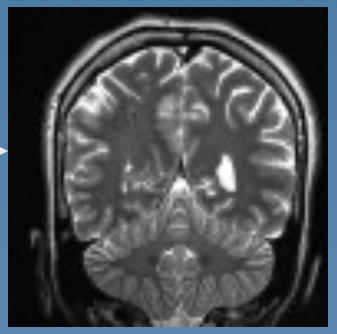


Image processing

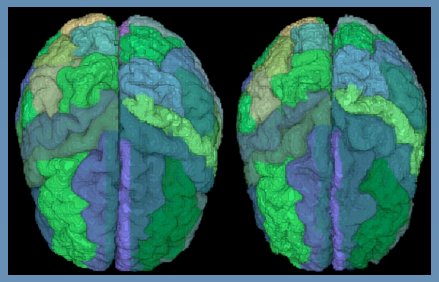
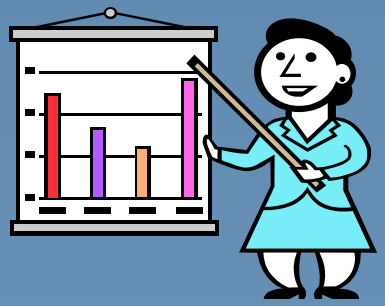


Image analysis

123……
2346……..
65789……
6578……..

Data output

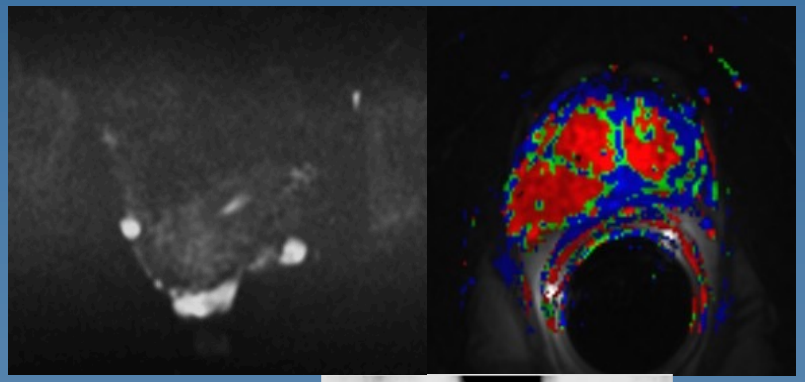


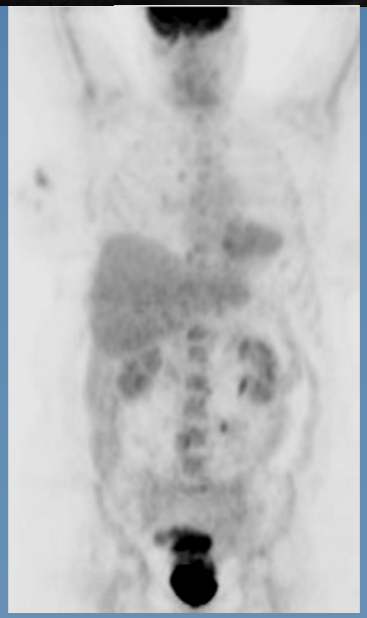
Data analysis

Clinical Trials: Imaging is an "Assay"

What is a good assay?

- Stable technology
- Available widely
- Standardized image acquisition
- Reproducible
- Range of normal defined





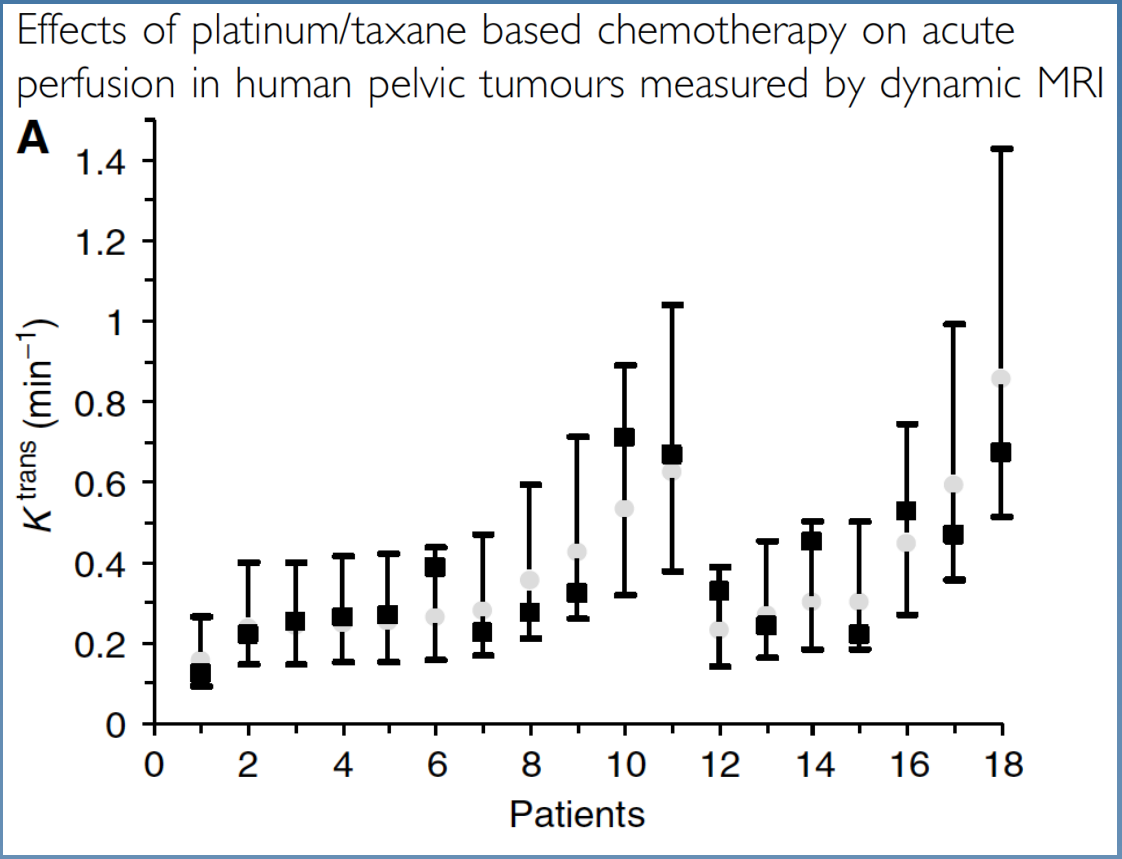
Balance "state of the art" with "generalizability"

Sargent DJ, Rubinstein L, Schwartz L et al. *Eur J Cancer* 2009; 45: 290.

Signal Requires Data Quality

Variability: Test – Retest

Same patient, day, scanning protocol but separate imaging sessions



Pre-treatment

Post-treatment

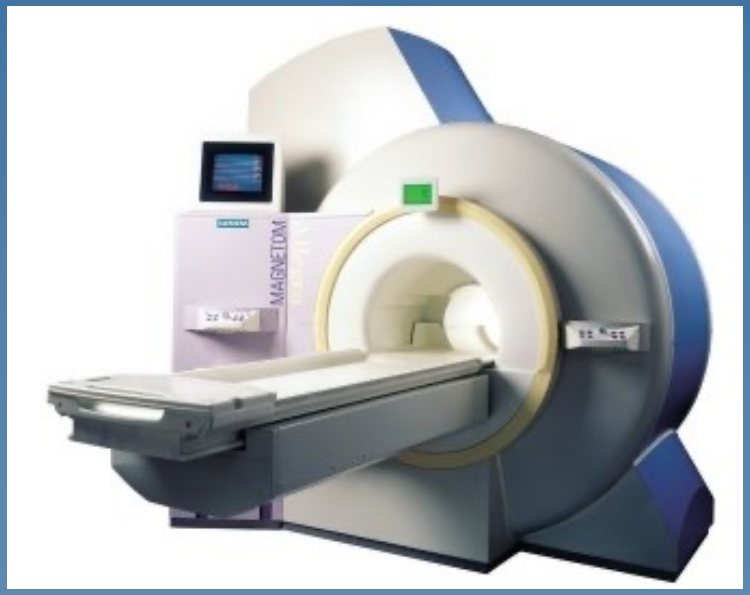
Range of test-retest

Conclusion

Index test variability precludes detecting pre- vs. post-treatment change.

Lankester KJ, Taylor NJ, Stirling JJ et al. Br J Cancer. 2005.93:979.

Imaging Chain: Patient → Data



Raw data

Image reconstruction

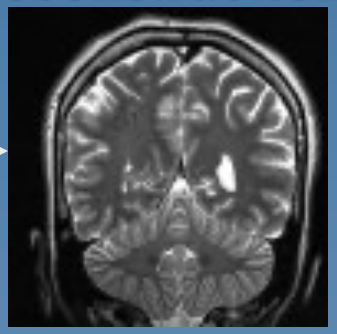


Image processing

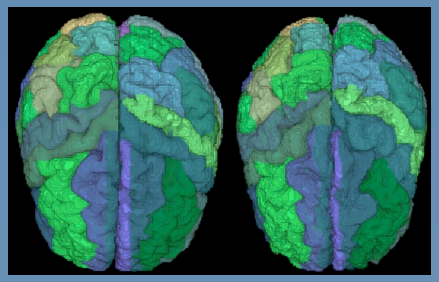
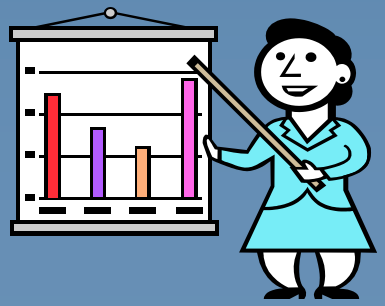


Image analysis

123……
2346……..
65789……
6578……..

Data output



Data analysis

Imaging Manual

- Hardware and software
- Scanner calibration
- Patient preparation
- Scanning protocol
- Post-processing

Imaging manual with a series of step by step SOP's (standard operating procedures) for image acquisition

CTMW Template: Imaging/Device Manual

13 APPENDIX III: NOVEL IMAGING OR NOVEL DEVICE MANUAL

No text needed here

13.1 IMAGING MANUAL

Imaging manual is intended to be used as a step-by-step description of the hardware, image acquisition, reader study and quality assurance for novel imaging. Content varies with the novel imaging to be performed but possible applicable content includes:

- *Scanner specifications*
- *Scanner qualification procedures*
- *Novel tracer – preparation, dose*
- *Image acquisition protocol for each scanner type*
  - *Radiation dose*
  - *Contrast dose*
- *Ongoing site quality control procedures (dose monitoring, calibration, etc.)*
- *Reader study description – number of readers, images reviewed, blinding, etc.*
- *Reader forms*

In general, standard-of-care clinical imaging for participant followup does not need to be described in the Imaging Manual.

<Insert text>

13.2 DEVICE MANUAL

Medical Device is defined, in part, as any health care product that does not achieve its primary intended purpose by chemical action or by being metabolized.

Provide the following Device Information:

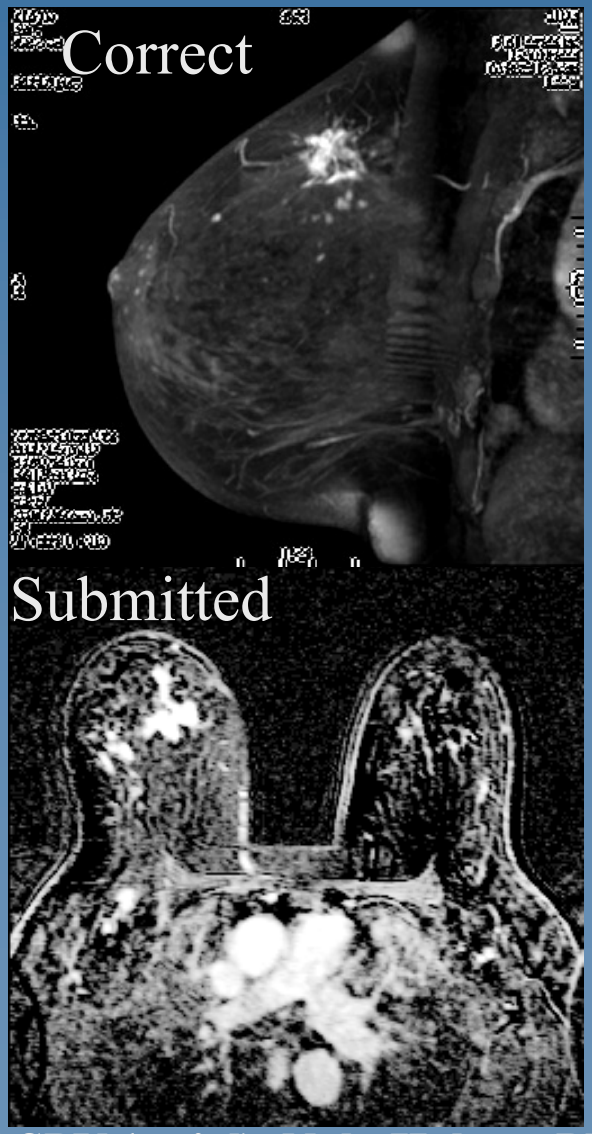
- *Device Name*
  - *Intended Use*
  - *Sponsor*
  - *Name*

Sources of Variability

- Patient preparation (e.g. fasting, serum lab values, tracer/contrast administration)
- Sampling (e.g. matrix, detector size, slice thickness)
- Reconstruction (e.g. filtering)

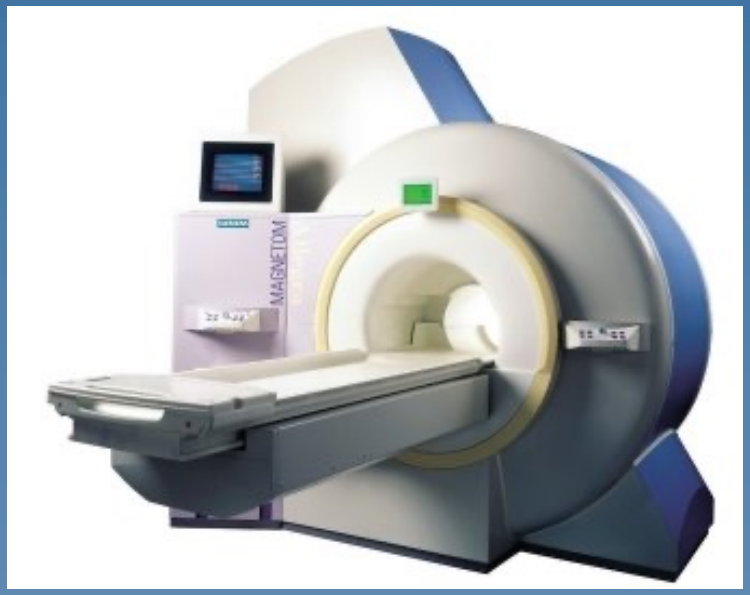
Ongoing QA of Incoming Data

- Data forms
  - Missing fields
  - Misinterpretation
- Images
  - Scanner performance
  - Acquisition protocol
- Regularly ”clean” your data
- Consider building in interim analysis up front



ACRIN 6657 (I-SPY 1 trial), Nola Hylton, PI

Imaging Chain: Patient → Data



Raw data

Image reconstruction

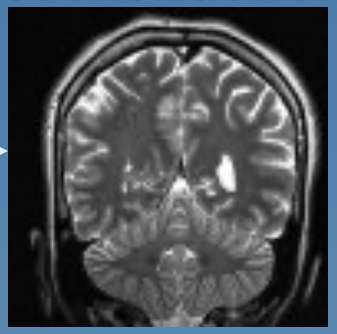


Image processing

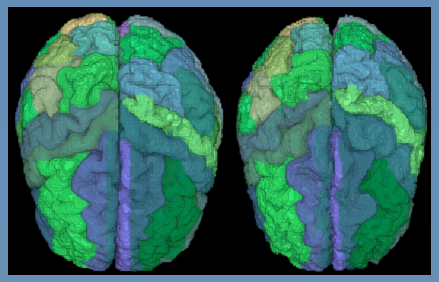
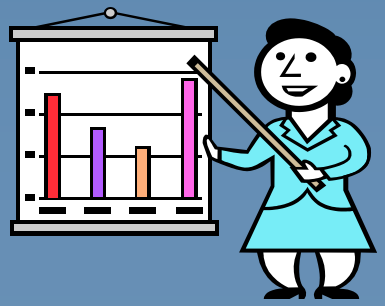


Image analysis

123……
2346……..
65789……
6578……..

Data output

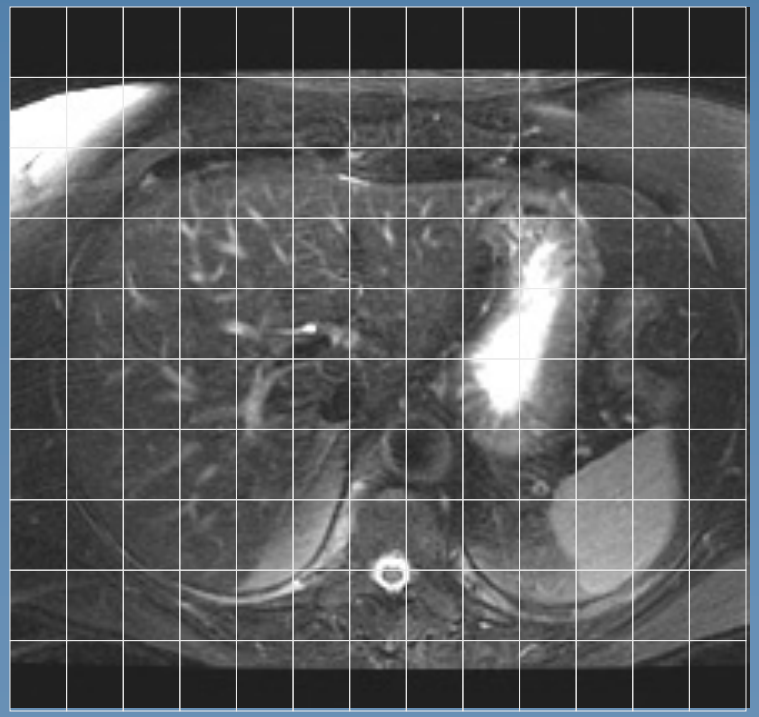


Data analysis

Image analysis: Turning image into data

- Reader extracted features
- Semi automated
- Fully automated

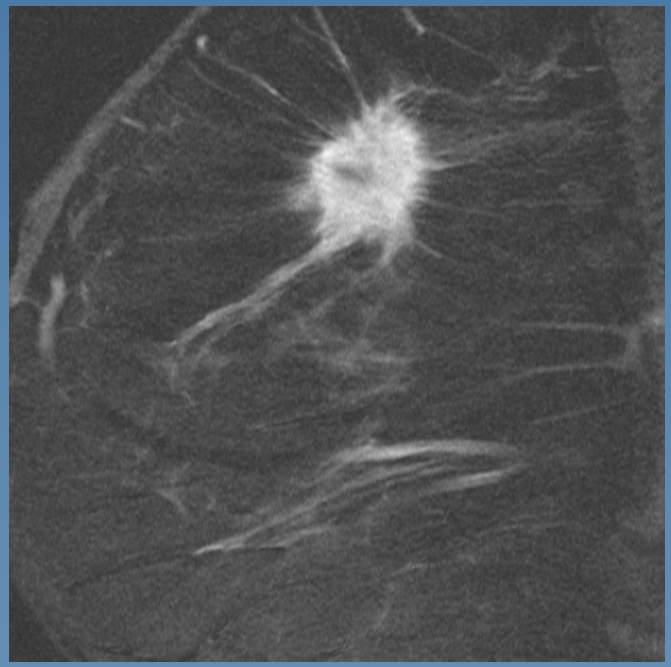
Feature 1
Feature 2
Feature 3
.
.
.



Reader Extracted Features

- Location
- Number
- Density/intensity/tracer uptake
- Shape (e.g. round, oval)
- Size (2D vs. 3D)
- Margin
- Enhancement kinetics/pattern
- Summary assessment

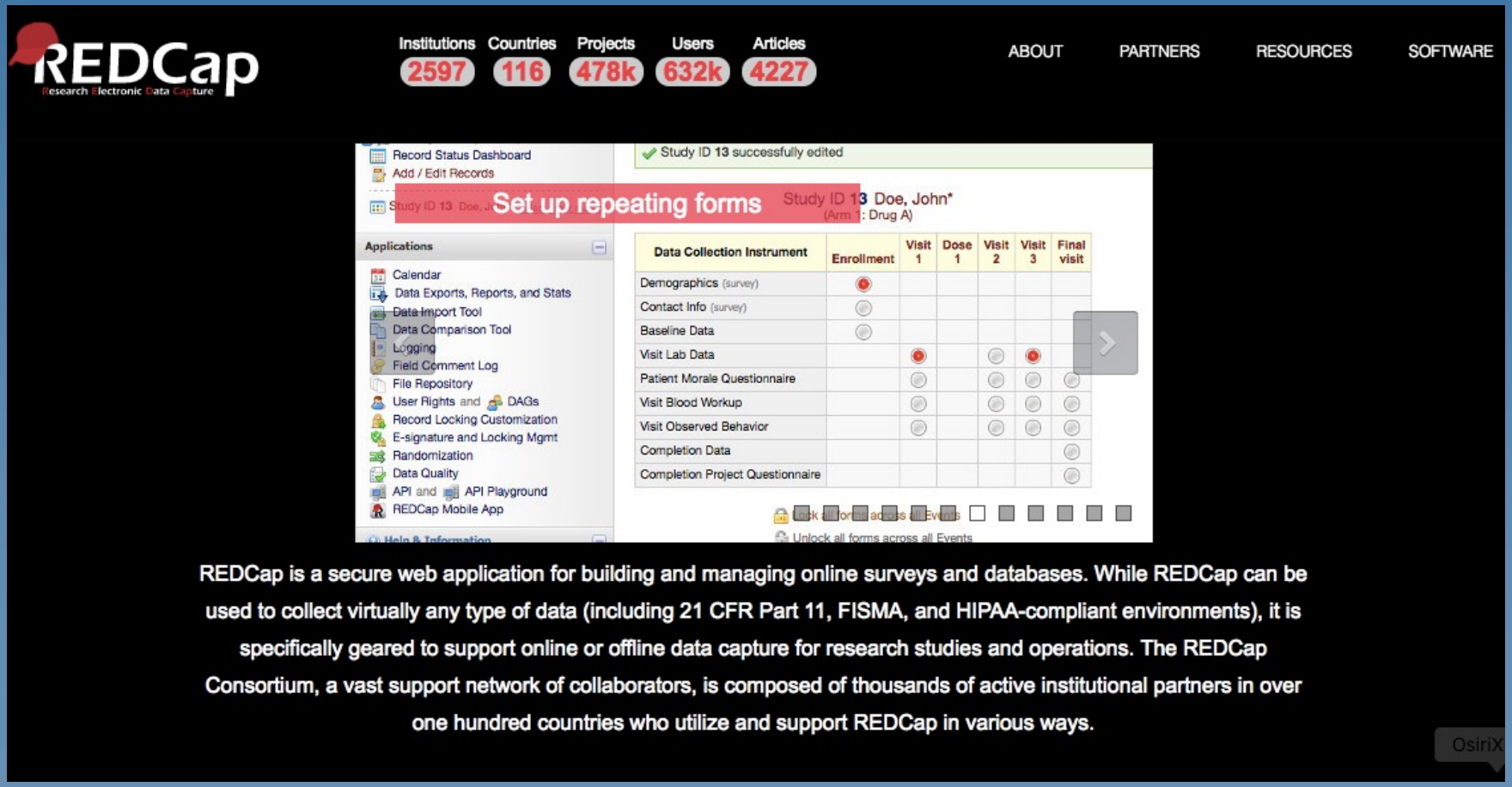
  "RADS"



Data Collection, Archiving & Sharing

- Will I need to collect this ?
  - Primary and secondary analyses
  - Planning the next phase trial
  - Compliance – IRB, HIPAA, FDA, sponsor, ICJME
- Is it essential that I collect this? – cost, time, and incomplete datasets
- Reports and retrieval
  - Database structure, search engine, image saves, etc.
- Policies on data sharing (ICJME requirement)

Cloud-Based Data Forms: REDCap



https://projectredcap.org/

REDCap Quality Module

REDCap Test Project

Data Quality

This module will allow you execute data quality rules upon your project data to check for discrepancies in your data. Listed below are some pre-defined data rules that you may utilize and run. You may also create your own rules or edit, delete, or reorder the rules you have already created. To find discrepancies for a given rule, simply click the Execute button next to it, or click the Execute All Rules button to fire all the rules at once. It will provide you with a total number of discrepancies found for each rule and will allow you to view the details of those discrepancies by clicking the View link next to each. Read more detailed instructions.

Data Quality Rules — Processing Complete! — Execute All Rules — Clear

Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Total Discrepancies	Site 2	Site 1
a	Missing values*	-	10,000+ view	1191	0
b	Missing values* (required fields only)	-	563 view	55	0
c	Field validation errors (incorrect data type)	-	26 view	6	0
d	Field validation errors (out of range)	-	18 view	2	0
e	Outliers for numerical fields (numbers, integers, sliders, calc fields)	-	25 view	1	0
f	Hidden fields that contain values**	-	28 view	4	0
g	Multiple choice fields with invalid values	-	17 view	6	0

Archiving Image Databases

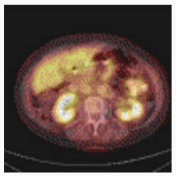
The Cancer Genome Atlas - Lung Adenocarcinoma PUBLIC

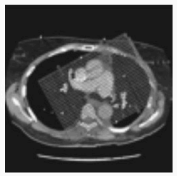
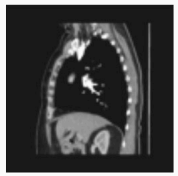
CT PET Chest

From The Cancer Imaging Archive (TCIA): the Cancer Genome Atlas Lung Adenocarcinoma data collection is part of a larger effort to buil d a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA). Clinical, genetic, and pathological data resides in the Genomic Data Commons (GDC) Data Portal while the radiological data is stored on The Cancer Imaging Archive.

DATASET	DATASET TYPE	# STUDIES	# SERIES	# IMAGES
Dataset	DICOM	152	625	48,931

SERIES THUMBNAILS



+ 621 more...

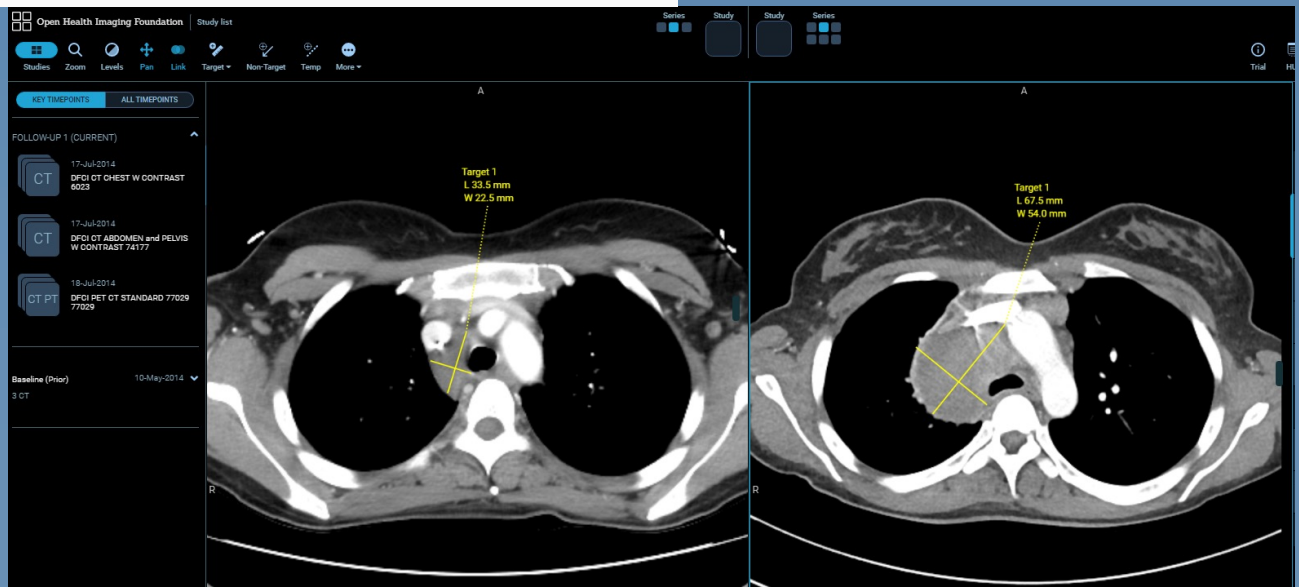
Users

MD.ai A G

Labels

Nodule

Emphysema

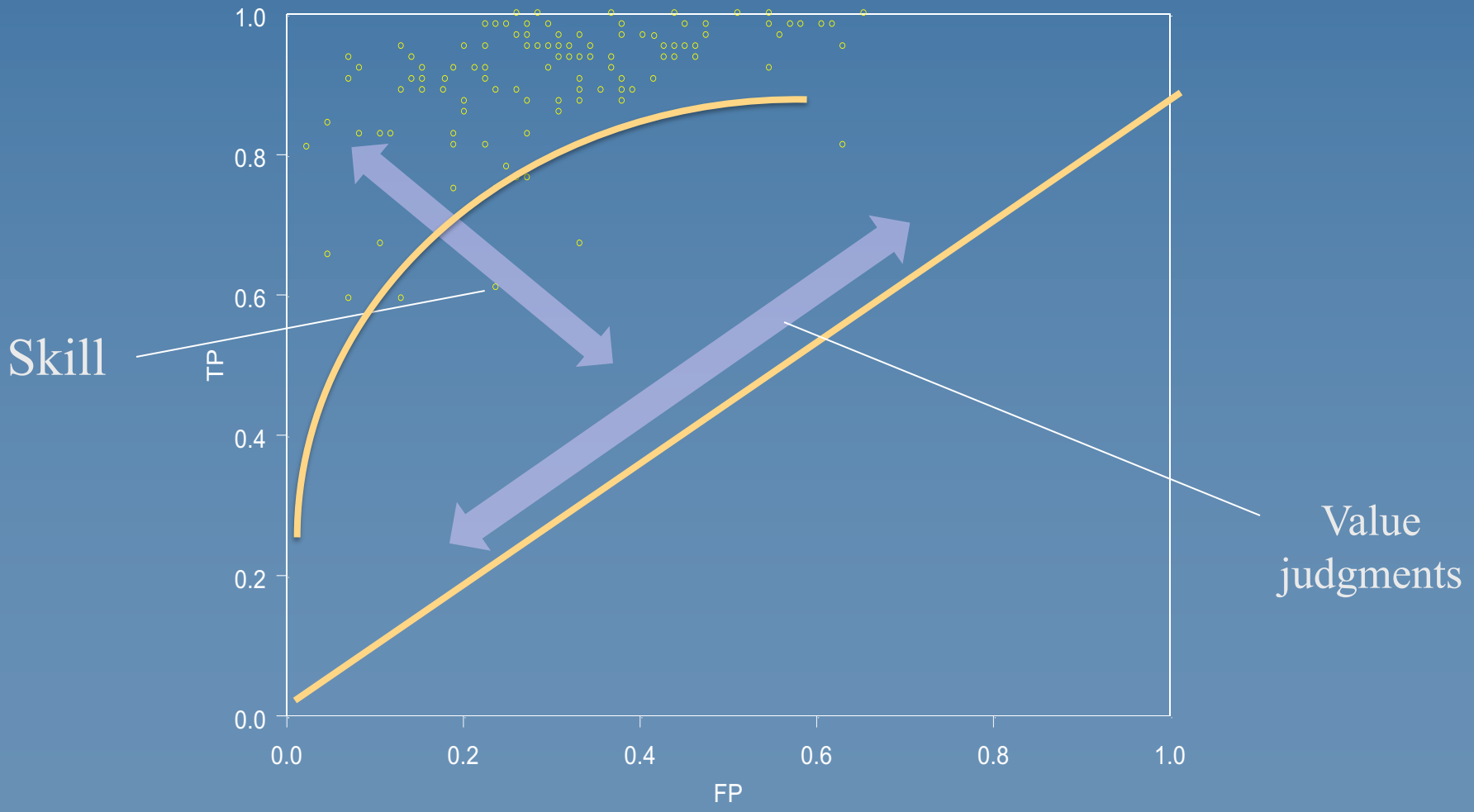


Reader Study

- Multiple independent readers
- Rules for image interpretation
  - Information available to reader
  - Image selection, windowing, order, etc.
  - Choosing index lesions
  - Selecting region of interest (ROI)
  - Definition of positive vs. negative test
- Reader training
- Digital data forms and time stamps for signoff

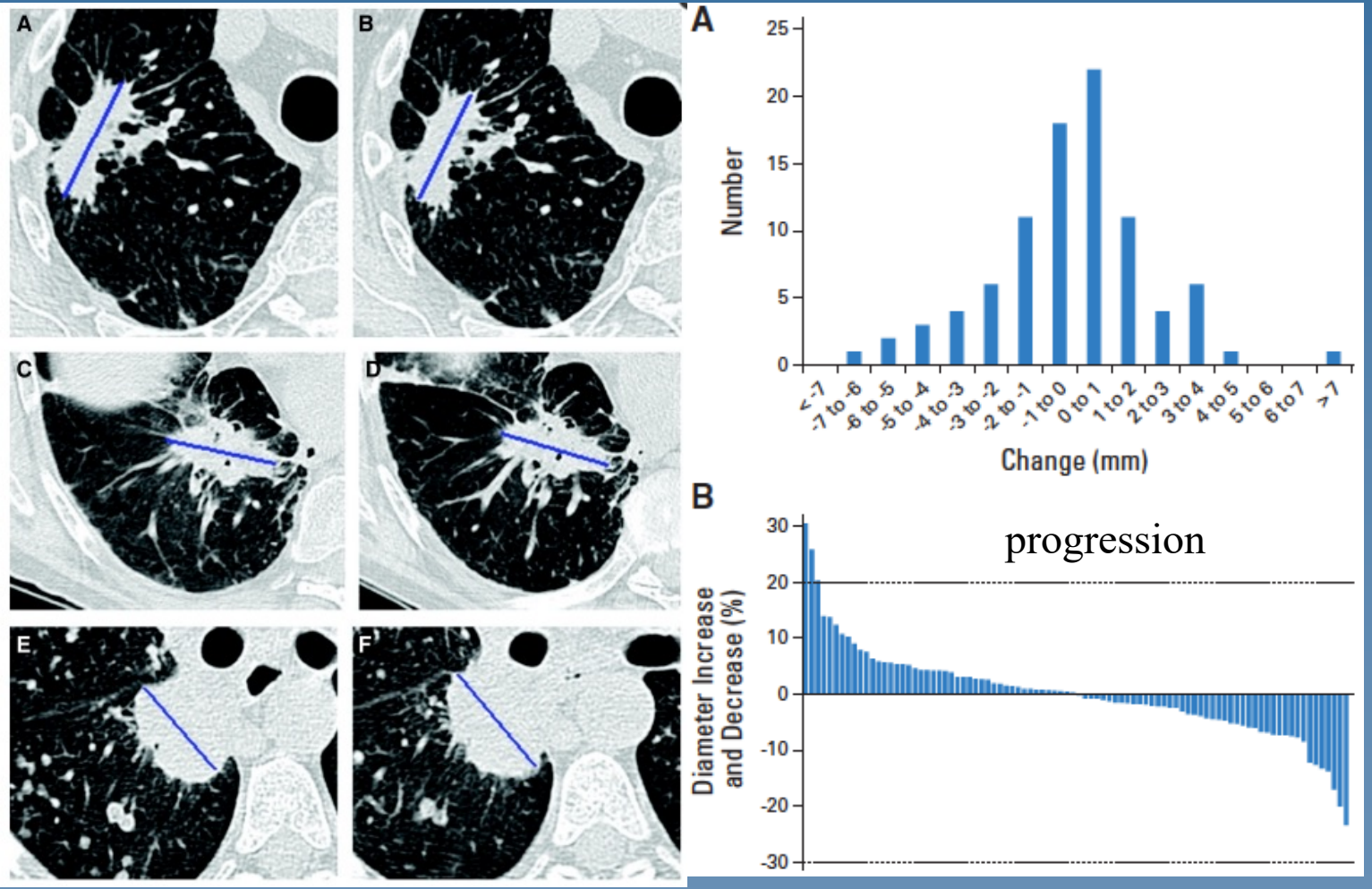
A manual defining reader rules and training cases are part of any prospective study design.

ROC operating points of 108 radiologists reading same mammograms



Beam, Layde, Sullivan Arch Intern Med 1996; 156:209-213

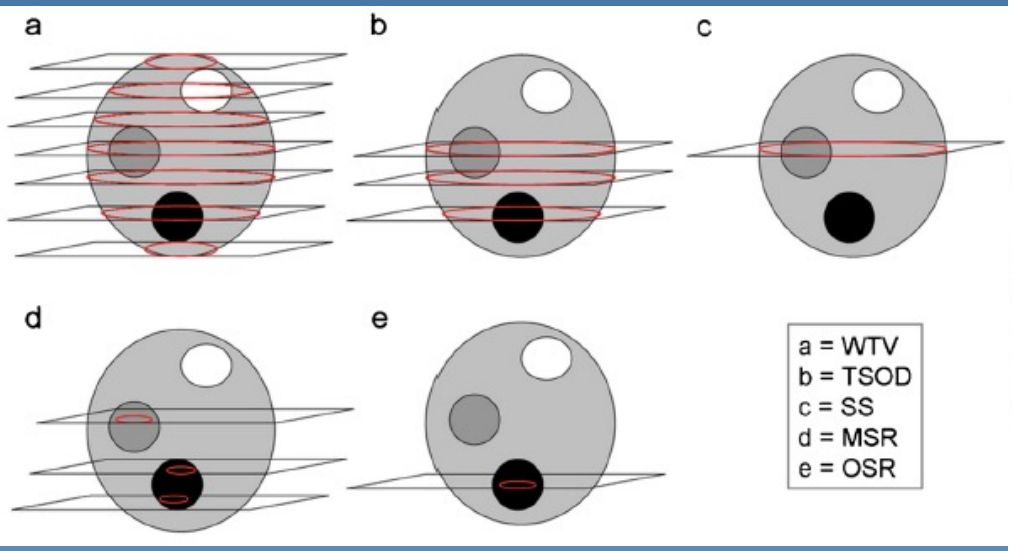
Reader Variability: Size

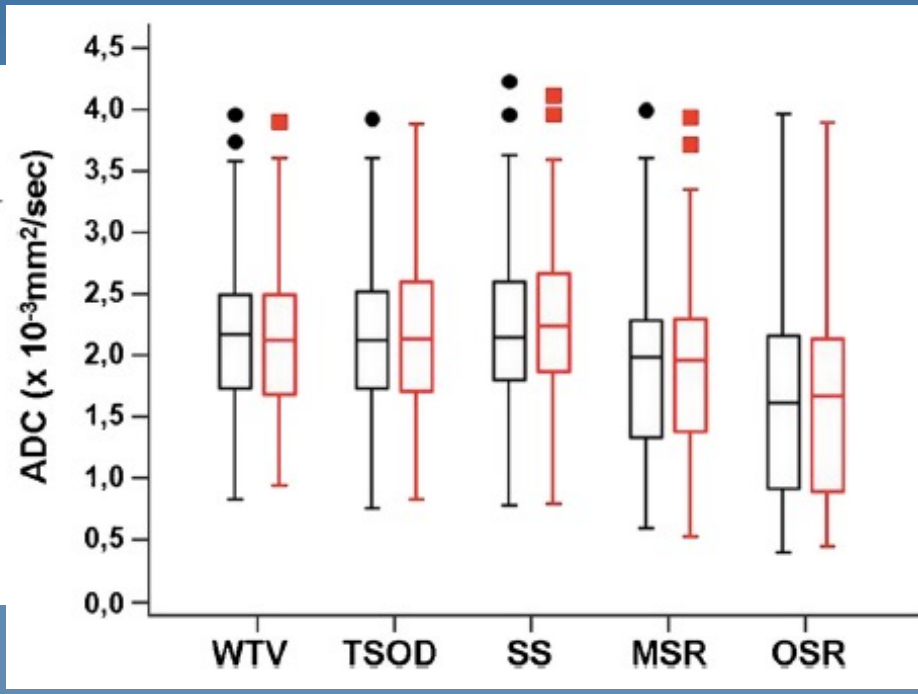


- Increases and decreases of <10% can be a result of inherent variability.

Oxnard GR et al. J Clin Oncol. 2011;29:3114-9.

Variability Introduced by ROI Selection





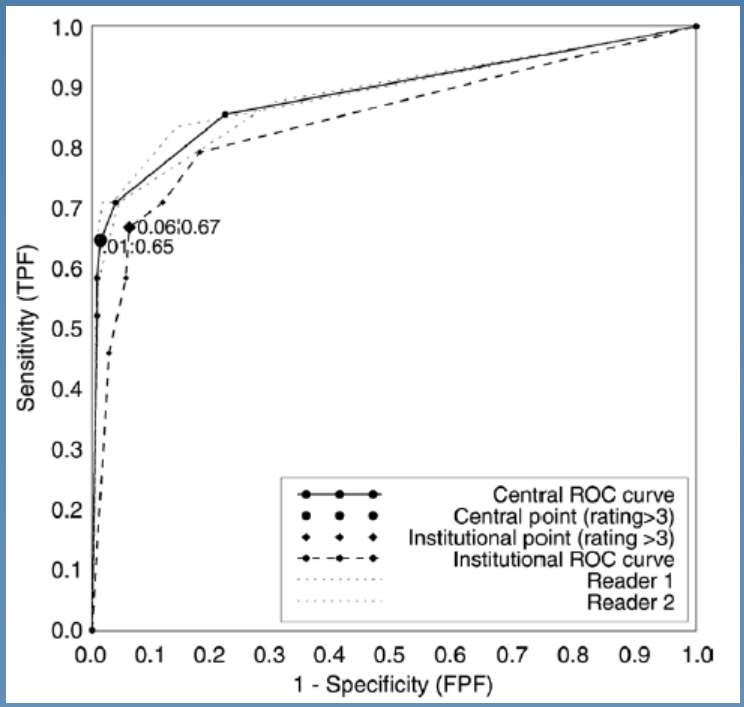
- All ROI protocols show excellent inter-observer agreement (ICC 0.94)
- Different ROI protocols yield different ADC values

Priola AM. . . Veltri A. Eur Radiol. 2017;27:1386-1394.

Site vs Central Reader ACRIN 6671: FDG-PET/CT Endometrial Cancer

Detection of Stage IV disease

	Central	Site
Sensitivity (95% CI)	65%(48, 79)	67% (45, 85)
Specificity (95% CI)	98% (97, 99)	94% (90, 97)



Gee MS, Atri M. . . Lee SI. Radiology. Nov 29, 2017:Epub.

AI-Assisted Readers

Radiology

ORIGINAL RESEARCH • MUSCULOSKELETAL IMAGING

Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence

Ali Guermazi, MD, PhD • Chadi Tannoury, MD • Andrew J. Kompel, MD • Akira M. Murakami, MD • Alexis Ducarouge, MSc • André Gillibert, MD, MSc • Xinning Li, MD • Antoine Tournier, MSc • Youmna Lahoud, MD • Mohamed Jarraya, MD • Elise Lacave, MSc • Hamza Rahimi, MD • Aloïs Pourchot, MSc • Robert L. Parisien, MD • Alexander C. Merritt, MD • Douglas Comeau, DO • Nor-Eddine Regnard, MD, MSc • Daichi Hayashi, MD, PhD

Radiology

ORIGINAL RESEARCH • THORACIC IMAGING

Content-based Image Retrieval by Using Deep Learning for Interstitial Lung Disease Diagnosis with Chest CT

Jooae Choe, MD, PhD • Hye Jeon Hwang, MD, PhD • Joon Beom Seo, MD, PhD • Sang Min Lee, MD, PhD • Jihye Yun, PhD • Min-Ju Kim, MSc • Jewon Jeong, MD • Youngsoo Lee, MD • Kiok Jin, MD • Rohee Park, MD • Jihoon Kim, MD • Howook Jeon, MD • Namkug Kim, PhD • Jaeyoun Yi, PhD • Donghoon Yu, MSc • Byeongsoo Kim, BSc

ORIGINAL RESEARCH • PEDIATRIC IMAGING

Radiology

Artificial Intelligence Algorithm Improves Radiologist Performance in Skeletal Age Assessment: A Prospective Multicenter Randomized Controlled Trial

David K. Eng, MS • Nishith B. Khandwala, MS • Jin Long, PhD • Nancy R. Fefferman, MD • Shailee V. Lala, MD • Naomi A. Strubel, MD • Sarah S. Milla, MD • Ross W. Filice, MD • Susan E. Sharp, MD • Alexander J. Towbin, MD • Michael L. Francavilla, MD • Summer L. Kaplan, MD • Kirsten Ecklund, MD • Sanjay P. Prabhu, MD • Brian J. Dillon, MD • Brian M. Everist, MD • Christopher G. Anton, MD • Mark E. Bittman, MD • Rebecca Dennis, DO • David B. Larson, MD, MBA • Jayne M. Seekins, DO • Cicero T. Silva, MD • Arash R. Zandieh, MD • Curtis P. Langlotz, MD, PhD , • Matthew P. Lungren, MD, MPH • Safwan S. Halabi, MD

Fully Automated AI Readers

Radiology

ORIGINAL RESEARCH • GASTROINTESTINAL IMAGING

Population-Scale CT-based Body Composition Analysis of a Large Outpatient Population Using Deep Learning to Derive Age-, Sex-, and Race-specific Reference Curves

Kirti Magudia, MD, PhD • Christopher P. Bridge, DPhil • Camden P. Bay, PhD • Ana Babic, PhD • Florian J. Fintelmann, MD • Fabian M. Troschel, MD • Nityanand Miskin, MD • William C. Wrobel, MD • Lauren K. Brais, MPH • Katherine P. Andriole, MD, PhD • Brian M. Wolpin, MD, MPH • Michael H. Rosenthal, MD, PhD

Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Fully Automated Volumetric Breast Density Estimation from Digital Breast Tomosynthesis

Aimilia Gastounioti, PhD • Lauren Pantalone, MPH • Christopher G. Scott, MS • Eric A. Cohen, MS • Fang F. Wu, BS • Stacey J. Winham, PhD • Matthew R. Jensen, BS • Andrew D. A. Maidment, PhD • Celine M. Vachon, PhD • Emily F. Conant, MD • Despina Kontos, PhD

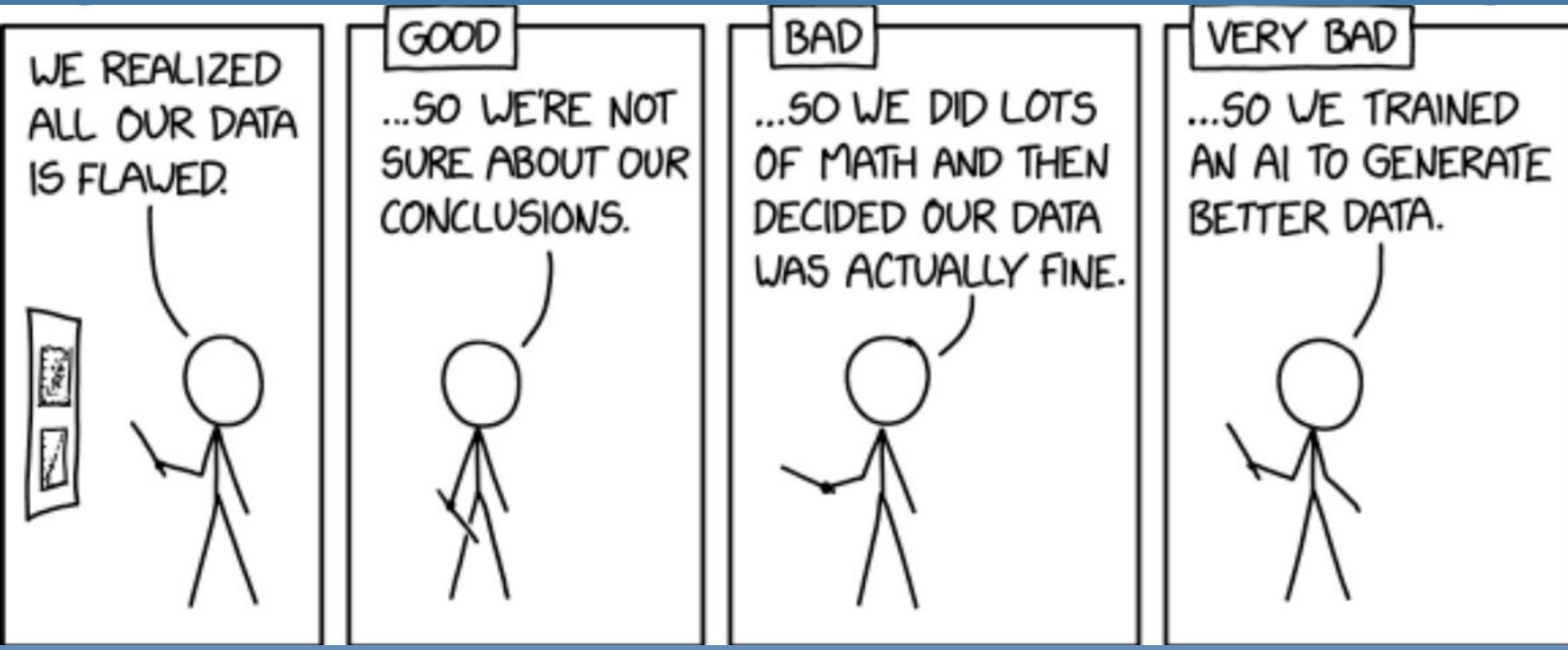
ORIGINAL RESEARCH • GASTROINTESTINAL IMAGING

Radiology

Opportunistic Osteoporosis Screening at Routine Abdominal and Thoracic CT: Normative L1 Trabecular Attenuation Values in More than 20 000 Adults

Samuel Jang, MD • Peter M. Graffy, BA, MPH • Timothy J. Ziemlewicz, MD • Scott J. Lee, MD • Ronald M. Summers, MD, PhD • Perry J. Pickhardt, MD

Inevitably we arrive back at AI. . .

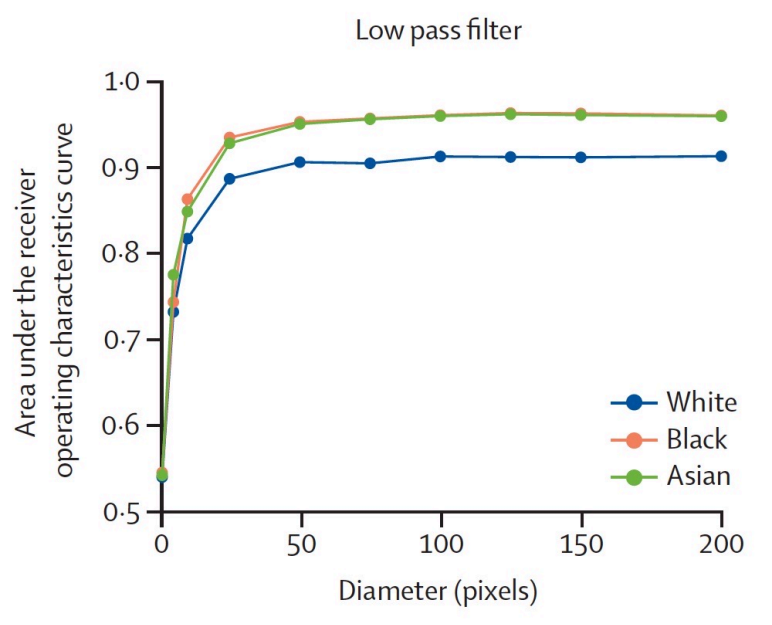


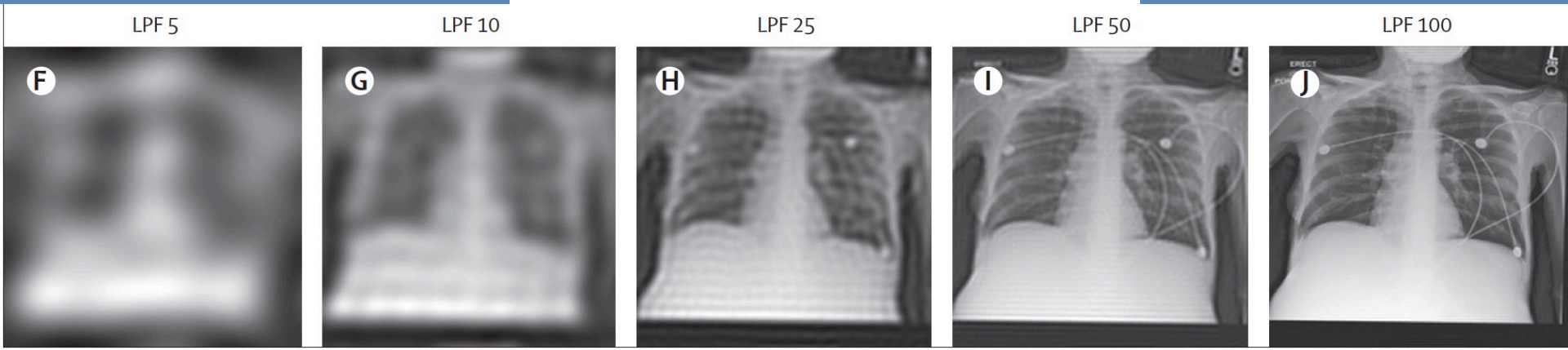
AI Database Landmines --- to date

Privacy: MRI & Facial Recognition

Schwarz CG. . . Jack CR . N Engl J Med. 2019;381:1684-1686.

Privacy: CXR & Race





Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA . . . .Trivedi H, Wang R, Zaiman Z, Zhang H. Lancet Digit Health. 2022;4(6):e406-e414.

Diversity and Bias: Cohorts Used to Train Deep Learning Algorithms

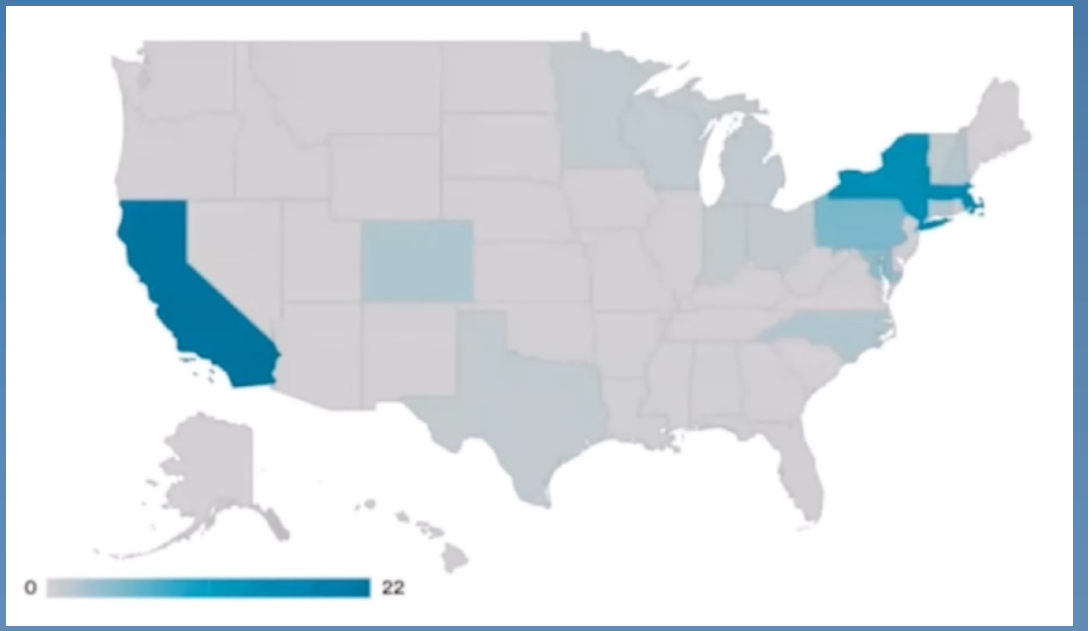
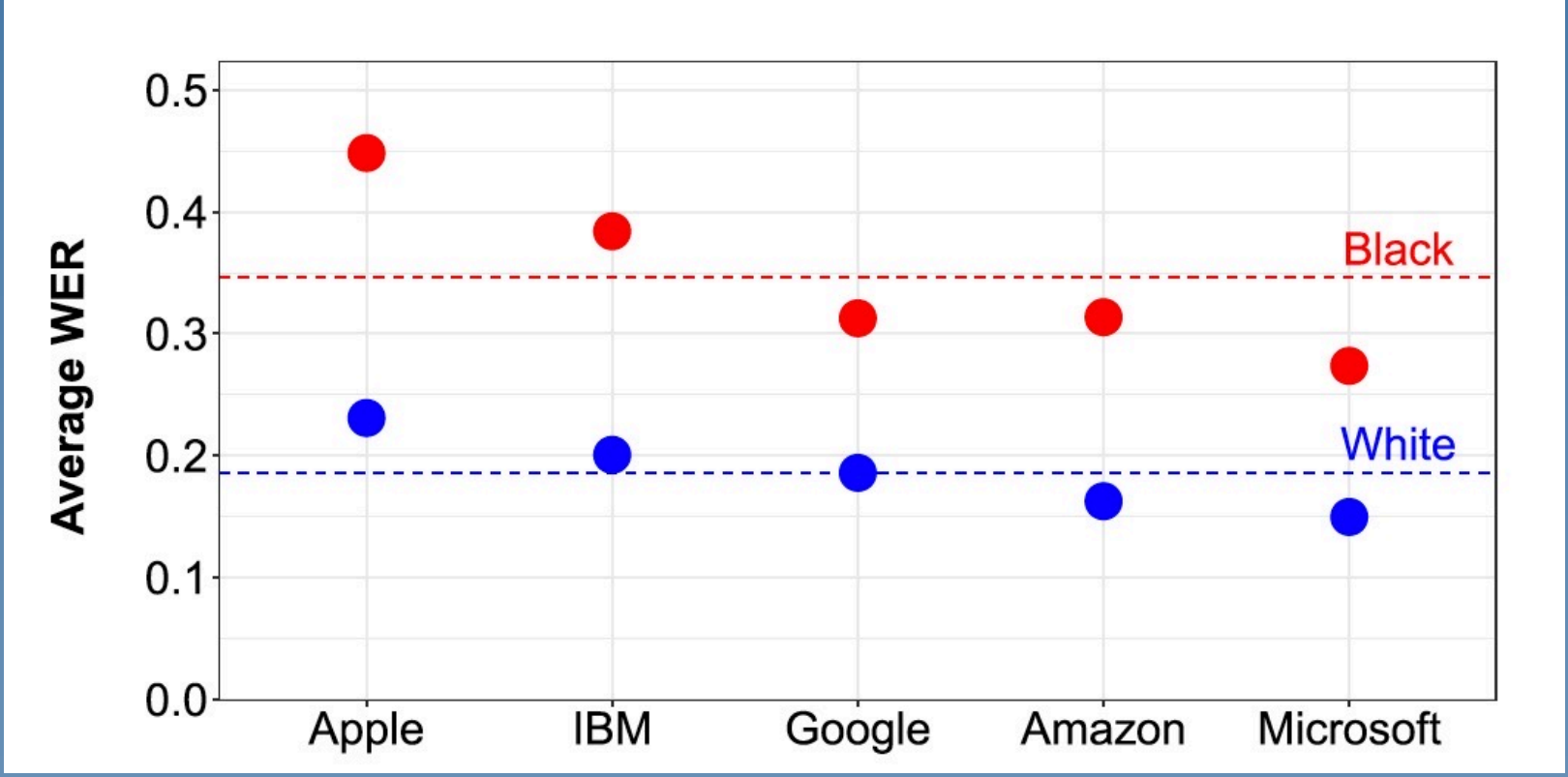


Table. US Patient Cohorts Used for Training Clinical Machine Learning Algorithms, by State[a]

States	No. of studies
California	22
Massachusetts	15
New York	14
Pennsylvania	5
Maryland	4
Colorado	2
Connecticut	2
New Hampshire	2
North Carolina	2
Indiana	1
Michigan	1
Minnesota	1
Ohio	1
Texas	1
Vermont	1
Wisconsin	1

Kaushal A, Altman R, Langlotz C. JAMA 2020;324:1212-1213.

Diversity and Bias: Racism Speech Recognition



Koenecke A, Nam A, Lake E, Nudell J, Quartey M, Mengesha Z, Toups C, Rickford JR, Jurafsky D, Goel S. Proc Natl Acad Sci U S A. 2020;117(14):7684-7689.

Steps toward data integrity:

- Image acquisition procedures
  - Balance "state of the art" with "generalizability"
  - Imaging manual
- Image analysis procedures
  - Read rules, training and testing, reader form
- Build in procedures for ongoing QA
- Database elements, architecture, and sharing policy --- with ethics in mind