Role of Ethics in AI for Healthcare

Melissa McCradden, PhD, MHSc
Bioethicist
The Hospital for Sick Children
Assistant Professor
University of Toronto, Canada

@mmccradden

Disclosures

Grant funding for research provided by:

- Canadian Institutes of Health Research
- Dalla Lana School of Public Health
- Edwin Leong Centre for Healthy Children
- SickKids Foundation
- AMS Healthcare
- National Institutes of Health (NIH, USA)
- National Health Services (NHSx, UK)

Talk Overview

- The ‘right’ kind of evidence
- Algorithmic bias and fairness
- Explainable AI?
- What makes a ‘good’ decision with AI?

‘DEPLOY’ AI IN HEALTHCARE

The right kind evaluation

Model performance *in silico*

Model in live data environment

We need technical validation + clinical evaluation

Responsible evaluation: more than just accuracy

'Clash of cultures': from technical validation to clinical evaluation and validation

Wongvibulsin & Zeger, 2020 *BMJ Evidence-Based Medicine*

Good technical performance does not always translate into patient benefit

- Advanced screening for certain cancers[1]
  - Increased healthcare spending, patient anxiety, low value
- 'Diagnostic downshift'?[2]
- Potential to increase workload without added value to patient care[3,4]
- Presumptions of low-risk may be ethically vulnerable[5,6]

[1]Shieh et al., 2016; [2]Sajid, Frost, & Paul 2021 *BMJ Ev Based Med*; [3]Tomašev et al., 2019 *Nat*; [4]Connell et al., 2019 *J Med Internet Res;* [5]Wilson et al., 2021 *BMJ*; [6]Wilson 2021 *BMJ*;

comment

Check for updates

Clinical research underlies ethical integration of healthcare artificial intelligence

Familiar concepts from research ethics can guide the meaningful and rigorous translation of artificial intelligence (AI) tools into clinical practice.

Melissa D. McCradden, Elizabeth A. Stephenson and James A. Anderson

NATURE MEDICINE | VOL 26 | SEPTEMBER 2020 | 1318–1330 | www.nature.com/naturemedicine

SickKids

Table 1 | Procedural and conceptual modifications for ethics review of healthcare ML research

Stage 1: data access	
Group-based approval	Allowing specified, qualified individuals grouped around common governance to access under defined terms and with a general goal
PHI protection	Removal of unnecessary forms of PHI while allowing the option of analyzing raw or masked data
Broad goal without pre-determined methodology	Avoids biasing research outputs and allows comparison of multiple methods to support implementation
Data-access frameworks	Stronger emphasis on the governance of data with accountability garnered through record-keeping of access and rationale
Pre-specified, frequent data retrieval without repeated amendments	Ensures model is learning from most current trends in health data
Stage 2: silent period	
Prospective non-interventional trial application as a template	Patients do not receive interventions; ML outputs do not reach the treating team to influence decision-making or the trial's evaluation
Goal of the trial	To test the hypothesis that the model is feasible and can have clinical applicability
Model validation	Calibration and technical performance assessed according to ML best practices
Clinical evaluation	Evidence generated for model's clinical applicability by comparison of silent predictions against real-time patient labels
Stage 3: clinical trial	
Goal of the trial	To test the hypothesis that the model is superior to the current standard of care
Generalizability	Aiming to demonstrate the generalizability of the approach, not the model itself
Disaggregated performance metrics	Disaggregated performance metrics are essential to patient safety and justice, as they will inform clinical uptake
Clinically relevant evaluation	To maximize clinical relevance and more precisely inform uptake, the plan for trial evaluation must include the following: • Model evaluated at its intended place in the decision-making pathway • Model outputs recorded • Clinical decisions recorded • Reasoning for discrepancy between output and decisions

PHI, protected health information.

THE AMERICAN JOURNAL OF BIOETHICS
https://doi.org/10.1080/15265161.2021.2013977

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

OPEN ACCESS Check for updates

SPIRIT-AI

CONSORT-AI

A Research Ethics Framework for the Clinical Translation of Healthcare Machine Learning

Melissa D McCradden[a,b,c], James A Anderson[a,d], Elizabeth A. Stephenson[e,f], Erik Drysdale[b], Lauren Erdman[b,g,h], Anna Goldenberg[a,g,h,i], and Randi Zlotnik Shaul[a,f,j]

Prospective Evaluation

Quasi-experimental trials

Quasi-experimental trials involve non-random participant selection, are practical, and can be used when ethical or logistic constraints prevent running a randomized trial. They are more systematic and less prone to bias than observational trials, but are more susceptible to confounding compared with RCTs.

Retrospective studies

Retrospective studies include model development and validation in an historical dataset. These studies offer some theoretical plausibility that a given label can be predicted using ML.

Exploratory

Randomized trials

The strongest evidence is provided by experimental studies (i.e., randomized controlled trials) which have the lowest risk of bias and offer greater confidence in establishing facts about the model's causal impact on patient outcomes.

Observational trials

Observational (non-interventional) trials are descriptive in nature and cannot establish facts about a model's impact on a given outcome. The silent trial is a form of observational study.

Silent trial

Figure 1. An evidentiary hierarchy for healthcare ML research.

SickKids

Responsible use of AI

- The current dominant approaches to validation are not well aligned with the informational needs of clinicians
- Considering ML models as components of an intervention ensemble may provide an empirical warrant for the judicious use of clinical AI to promote patient benefit*

The S·T·R·U·C·T·U·R·E of Clinical Translation:

Efficiency, Information, and Ethics

BY JONATHAN KIMMELMAN AND ALEX JOHN LONDON

The principal output of clinical translation is information—about the coordinated set of materials, practices, and constraints needed to safely unlock the therapeutic or preventive activities of drugs, biologics, and diagnostics.

*McCradden MD, Joshi S, Anderson JA, London AJ. *Work in progress*

Clinical Evidence

Responsible Translation:

- **Patient Benefit**
- **Contribution to medical knowledge system**
- **Reliability**
- **Aligning with informational needs**

Novel Interventions

Responsible translation – for all?

Fairness

RESEARCH

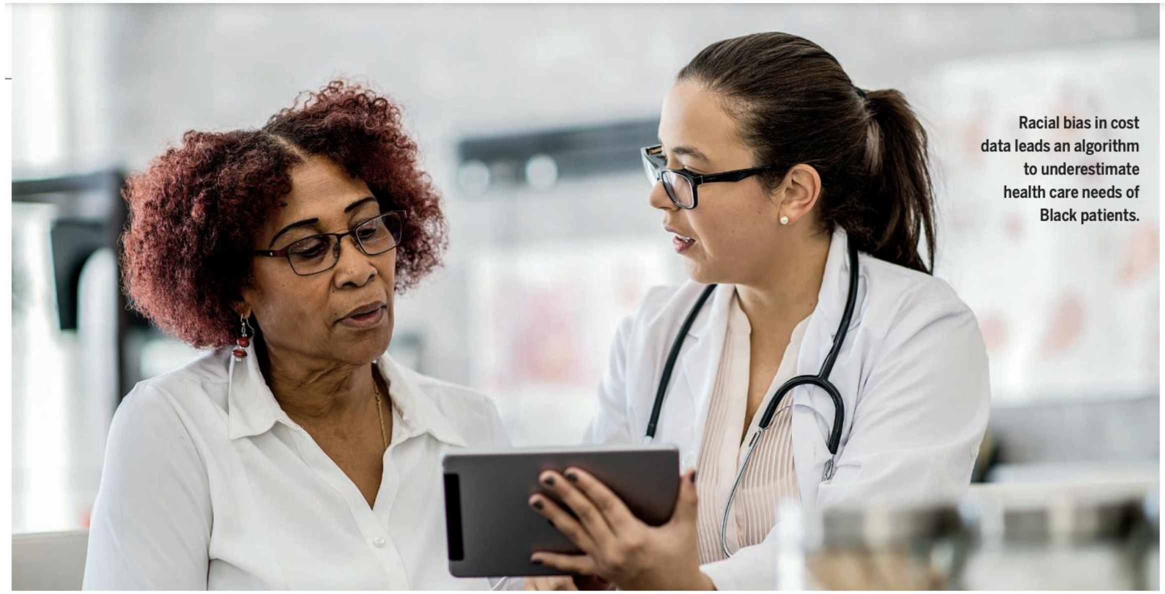
RESEARCH ARTICLE

ECONOMICS

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.



Racial bias in cost data leads an algorithm to underestimate health care needs of Black patients.

SOCIAL SCIENCE

Assessing risk, automating racism

A health care algorithm reflects underlying racial bias in society

Ruha Benjamin

+ See all authors and affiliations

Science 25 Oct 2019:
Vol. 366, Issue 6464, pp. 421-422
DOI: 10.1126/science.aaz3873

RESEARCH ARTICLE | HEALTH EQUITY

HEALTH AFFAIRS > VOL. 41, NO. 2: RACISM & HEALTH

Negative Patient Descriptors: Documenting Racial Bias In The Electronic Health Record

Michael Sun, Tomasz Oliwa, Monica E. Peek, and Elizabeth L. Tung

The Lancet Child & Adolescent Health

Volume 5, Issue 2, February 2021, Pages 103-112

Articles

Hospital outcomes for children with severe sepsis in the USA by race or ethnicity and insurance status: a population-based, retrospective cohort study

Hannah K Mitchell BMBS [a], Anireddy Reddy MD [b], Diana Montoya-Williams MD [c], Michael Harhay PhD [d, e, f], Jessica C Fowler MD [b], Nadir Yehya MD [b]

Gender bias concerns raised over GP app

Written by Sam Trendall on 13 September 2019 in Features

Onlookers are asking why the chatbot created by Babylon Health – which provides the GP at Hand service – is offering such different guidance to men and women. But the company tells *PublicTechnology* its service is working as intended.

SCIENTIFIC REPORTS

ticle | OPEN | Published: 15 April 2019

Genetic risk factors identified in populations of European descent do not improve the prediction of osteoporotic fracture and bone mineral density in Chinese populations

Yu-Mei Li, Cheng Peng, Ji-Gang Zhang, Wei Zhu, Chao Xu, Yong Lin, Xiao-Ying Fu, Qing Tian, Lei Zhang, Yang Xiang, Victor Sheng & Hong-Wen Deng

Scientific Reports **9**, Article number: 6086 (2019) | Download Citation

The Lancet Digital

Volume 3, Issue 4, April 2021, Pa

HEALTH

AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind

Machine learning has the potential to save thousands of people from skin cancer each year—while putting others at greater risk.

ANGELA LASHBROOK AUG 16, 2018

Articles

Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study

Rahuldeb Sarkar MPH [a, b, c], Christopher Martin PhD [d, e], Heather Mattie PhD [f], Judy Wawira Gichoya MD [g], David J Stone MD [h, i, j], Leo Anthony Celi PhD [f, k, l]

SickKids

Measuring Fairness in an Unfair World

Jonathan Herington
Department of Philosophy
University of Rochester
Rochester NY USA
Jonathan.Herington@rochester.edu

AIES'20, February 7–8, 2020, New York, NY, USA.

ACM 978-1-4503-7110-0/20/02…$15.00
DOI: https://doi.org/10.1145/3375627.3375854

Patterns

CellPress
OPEN ACCESS

Perspective

Algorithmic injustice: a relational ethics approach

Abeba Birhane[1,*]
[1]School of Computer Science, University College Dublin, Ireland & Lero—The Irish Software Research Centre, Dublin, Ireland
*Correspondence: abeba.birhane@ucdconnect.ie
https://doi.org/10.1016/j.patter.2021.100205

‘Algorithmic’ bias?

The NEW ENGLAND JOURNAL of MEDICINE

MEDICINE AND SOCIETY

Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms

Darshali A. Vyas, M.D., Leo G. Eisenstein, M.D., and David S. Jones, M.D., Ph.D.

Is poor performance a harm in itself?

- Though often conceptualized under **nonmaleficence**, discrepant algorithmic performance is not necessarily a harm *per se*
  - Centres the model performance in conceptualizing a harm → displaces ***people*** as the subjects of harm
  - E.g., in Obermeyer's (2019) paper they note that physicians likely redressed some of the observed algorithmic bias
- Some biases may not be correctable in the short-term
- What are some ways we can redistribute or redress these biases?
  - E.g., predicting no-shows

Can AI 'see' a patient's race?

- DL can "trivially detect" patient race based solely on image pixel data across an array of clinical tasks
  - No obvious reasons identified by the authors
- "A direct vector for the reproduction or exacerbation of the racial disparities that already exist in medical practice"

Experiments	AUC of Race Classification		
	Asian	Black	White
A1. Primary race detection in CXR Imaging			
MXR Resnet34	0.98	0.97	0.97
CXP Resnet34	0.97	0.98	0.97
EMX Resnet34	0.96	0.99	0.98

Table 3: Performance of deep learning models for the task of race detection on three large scale chest x-ray datasets. Values reflect the area under the receiver operating characteristic curve for each model on the test set (AUC).

Experiments	Asian	Black	White
A1. External validation of race detection models in CXR imaging			
MXR Resnet34 to CXP	0.93	0.97	0.93
MXR Resnet34 to EMX	0.89	0.97	0.95
CXP Resnet34 to MXR	0.97	0.97	0.96
CXP Resnet34 to EMX	0.89	0.96	0.91
EMX Resnet34 to MXR	0.96	0.98	0.97
EMX Resnet34 to CXP	0.95	0.98	0.95

Table 4: External validation performance of deep learning models for the task of race detection on three large scale chest x-ray datasets. Values reflect the area under the receiver operating characteristic curve for each model on the test set (AUC).

Banerjee et al., 2021 "Reading Race: AI Recognizes Patient's Racial Identity In Medical Images"
https://arxiv.org/pdf/2107.10356.pdf

Distributive Justice

- Are benefits and burdens distributed equally?
- Address the distribution? Or redress residual discrepancies?
- How we adjudicate between these approaches is fundamentally an ethical endeavour

“Diversity Hands” by Oswaldo Guayasamin

COMMENT | VOLUME 2, ISSUE 5, E221-E223, MAY 01, 2020

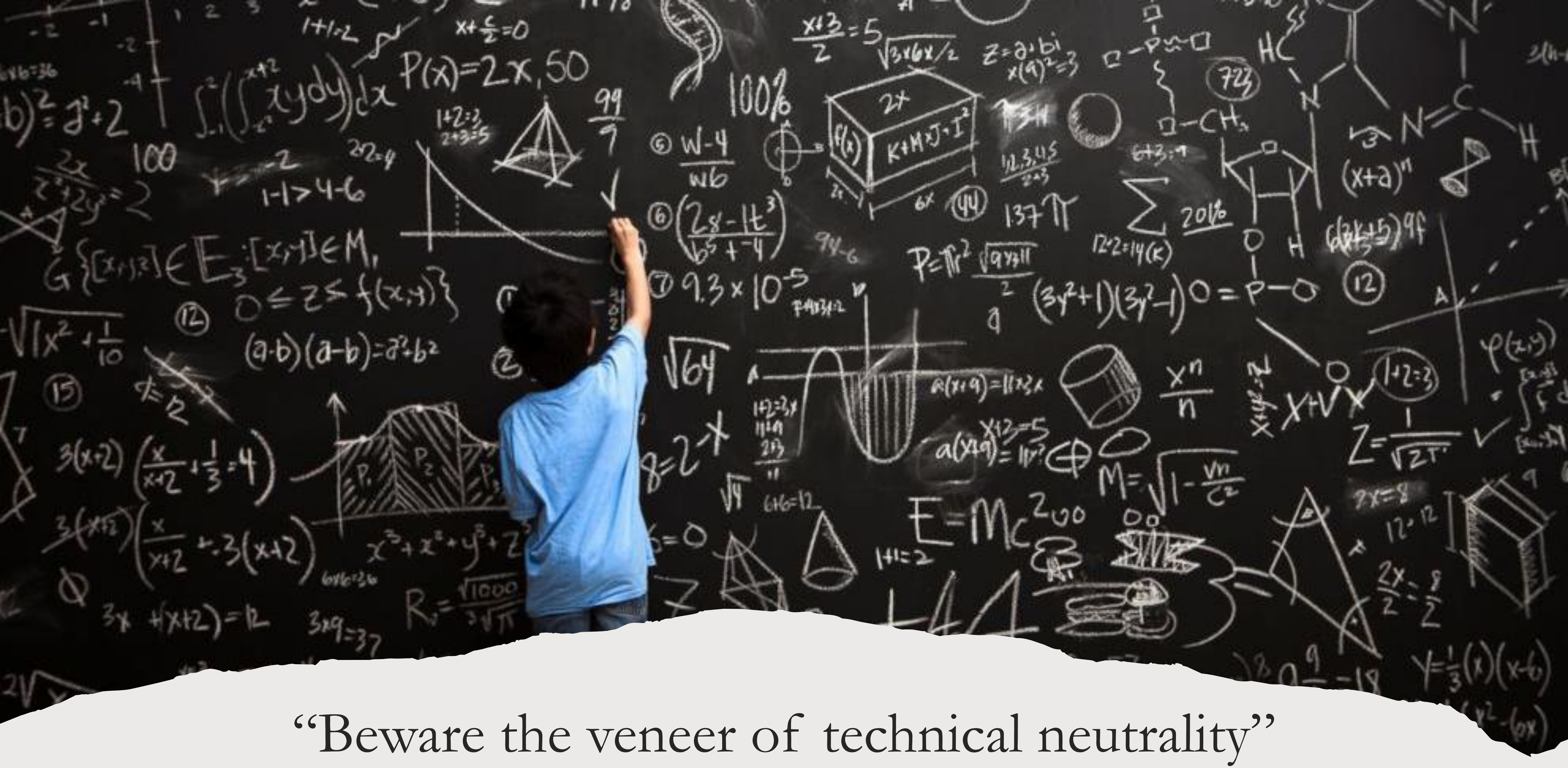
Ethical limitations of algorithmic fairness solutions in health care machine learning

Melissa D McCradden • Shalmali Joshi • Mjaye Mazwi • James A Anderson

Open Access • Published: May, 2020 • DOI: https://doi.org/10.1016/S2589-7500(20)30065-0

THE LANCET Digital Health

- Decisions about bias require two axes of consideration
  - Epistemic: what do we know about why these patterns are apparent?
  - Empirical: how will we evaluate the model's performance?
- Fairness is not simply <u>achieved</u> through model performance alone
  - E.g., referral parity vs outcome predictions
- Distributive Justice = characterize empirical performance & make ethical decisions for translation and clinical use



“Beware the veneer of technical neutrality”

Dr. Ruha Benjamin

Explainability

Is explainability the answer? Or does it raise more problems?

Explainable AI (XAI)

Generally, XAI focuses on helping the user ‘understand’:

- how the **model** works as a system
- how it arrived at a particular **prediction**

Ethical motivations for explainable AI

- Responsible AI-inclusive decision-making
- Informed consent and assent of patients and families
- Transparency

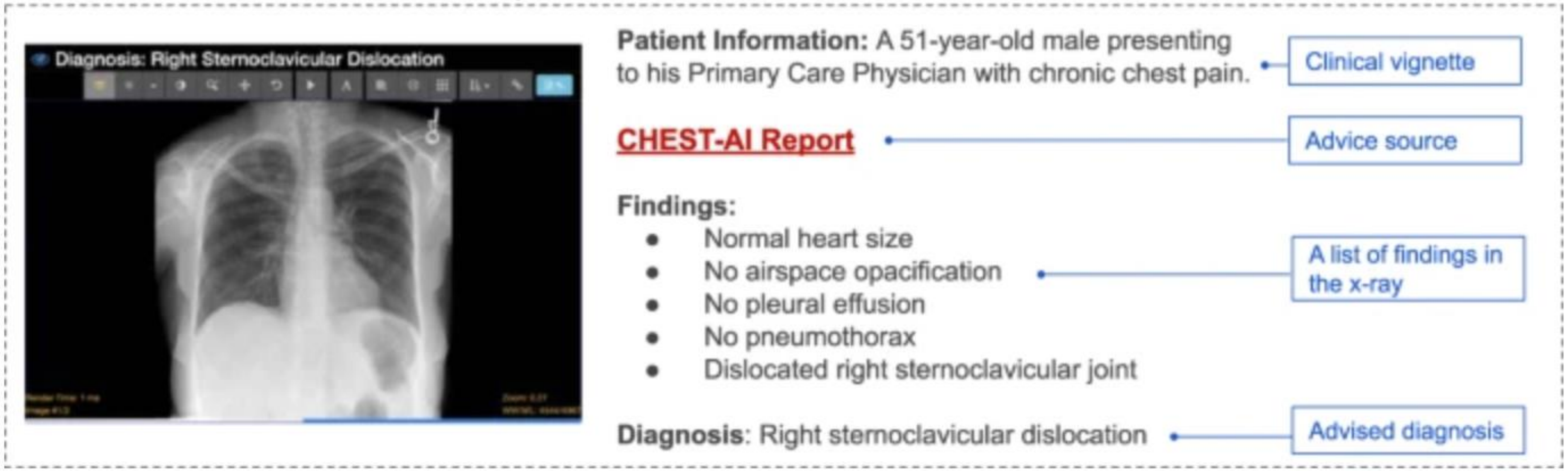
Does explainability actually achieve the ethical goals for which it is intended?
Does it pose other concerns?

Criticisms of XAI

- Computational explanations are sometimes not relevant for clinical decisions, do not reflection the metrics clinicians really care about[1,2]
- Many proposed explanations do not actually require machine learning, but are more about human-computer interaction or engineering[3]
- Concerns about **over-trust** (act on wrong outputs): user uncertainty, task complexity, and specific clinicians may be particularly likely to over-trust[4]
- These risks **are not restricted to 'black box'** systems alone
  - If opacity was the problem, explanations would prevent acting on wrong outputs; this is not what appears to be happening

[1]Lindsell et al. "Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem" 2020 *JAMA*; [2]Lipton "The Mythos of Model Interpretability" 2016 *ICML;* [3]Poursabzi-Sangdeh et al, "Manipulating and measuring model interpretability" 2021; [4]Ghassemi, "Don't Expl-AI-n Yourself" 2020 *C4E Talks*

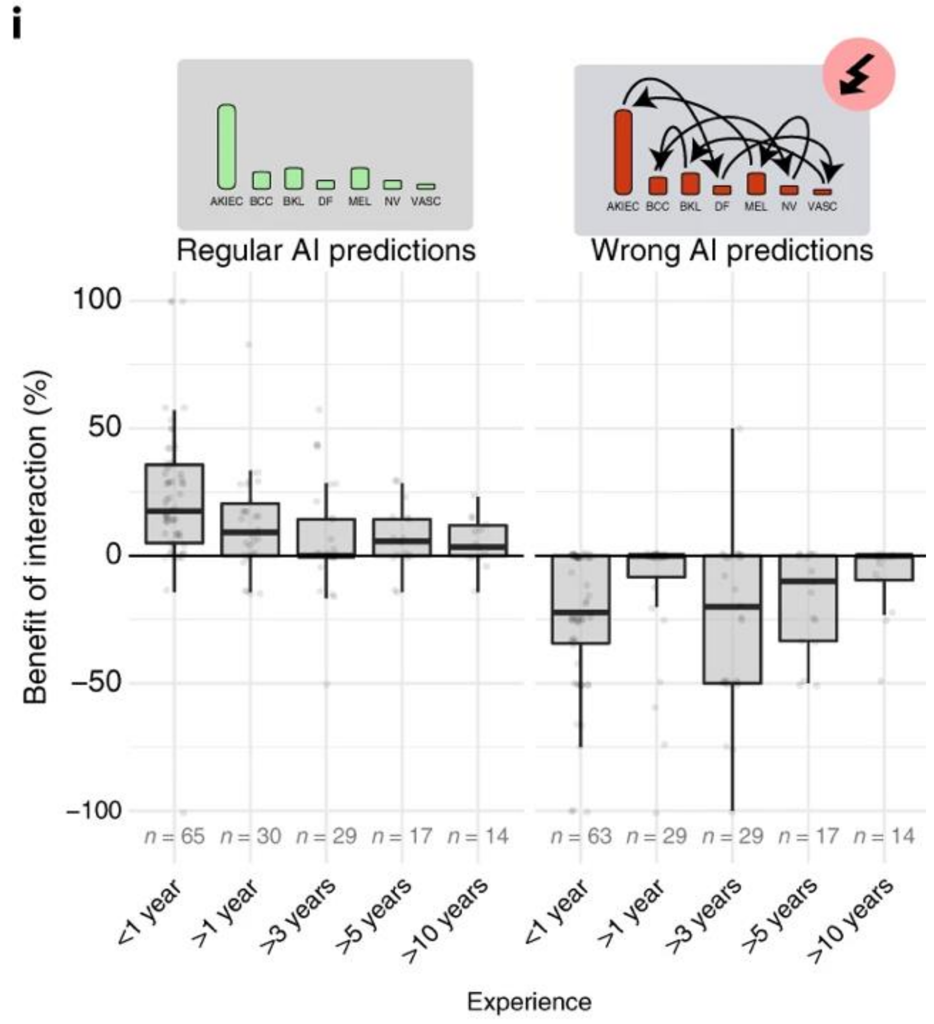
Case 1: Radiological reports



- Chest X-rays reports from CHEST-AI or Expert Radiologist
- Negative effect overall from receiving incorrect information (regardless of source), radiologists less so than internal/emergency medicine physicians
- Some clinicians are highly **susceptible to incorrect advice** while others are not

Case 2: Skin cancer recognition

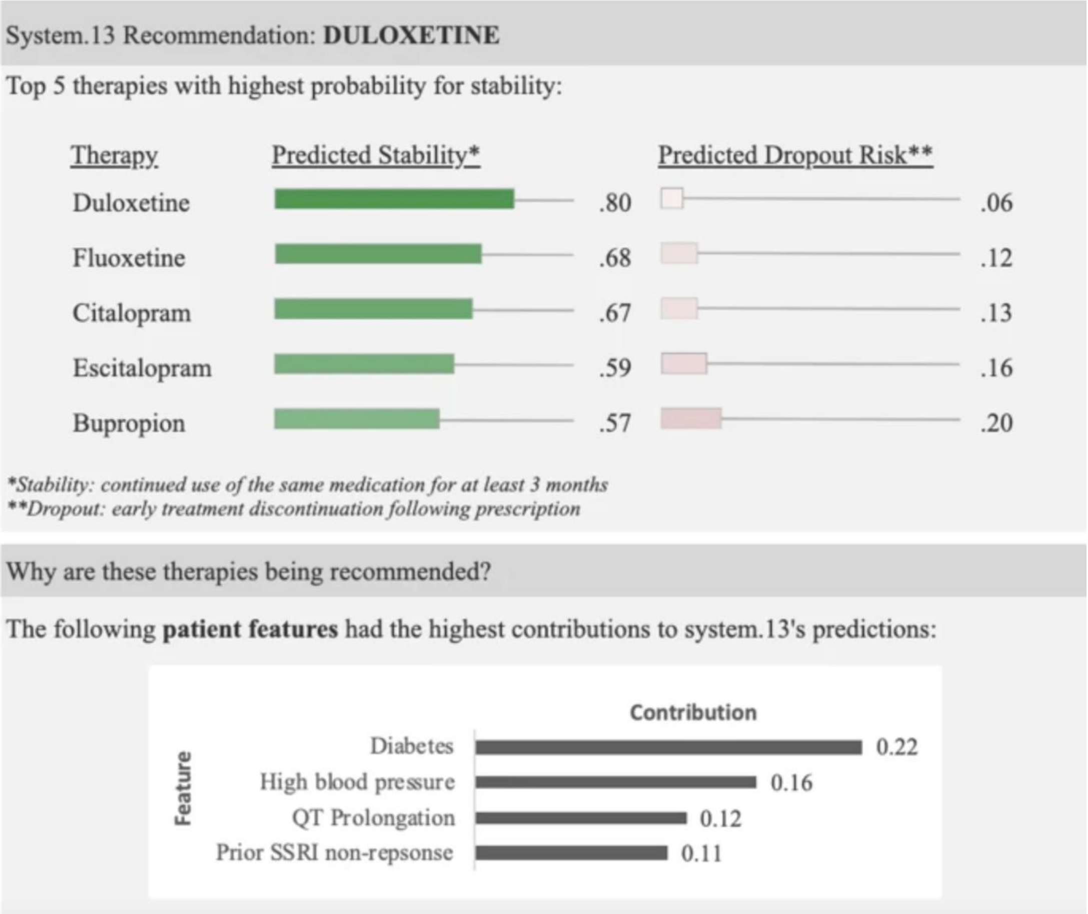
- Prediction of 7 distinct skin cancers
- Predictions = **correct** → clinician accuracy **improved**
- Predictions = **incorrect** → clinicians were often **misled**
- Changing one's mind was correlated with pre-prediction confidence
- Evident across the spectrum of clinical experience



Tschandl et al., "Human–computer collaboration for skin cancer recognition" 2020 *Nat Med*

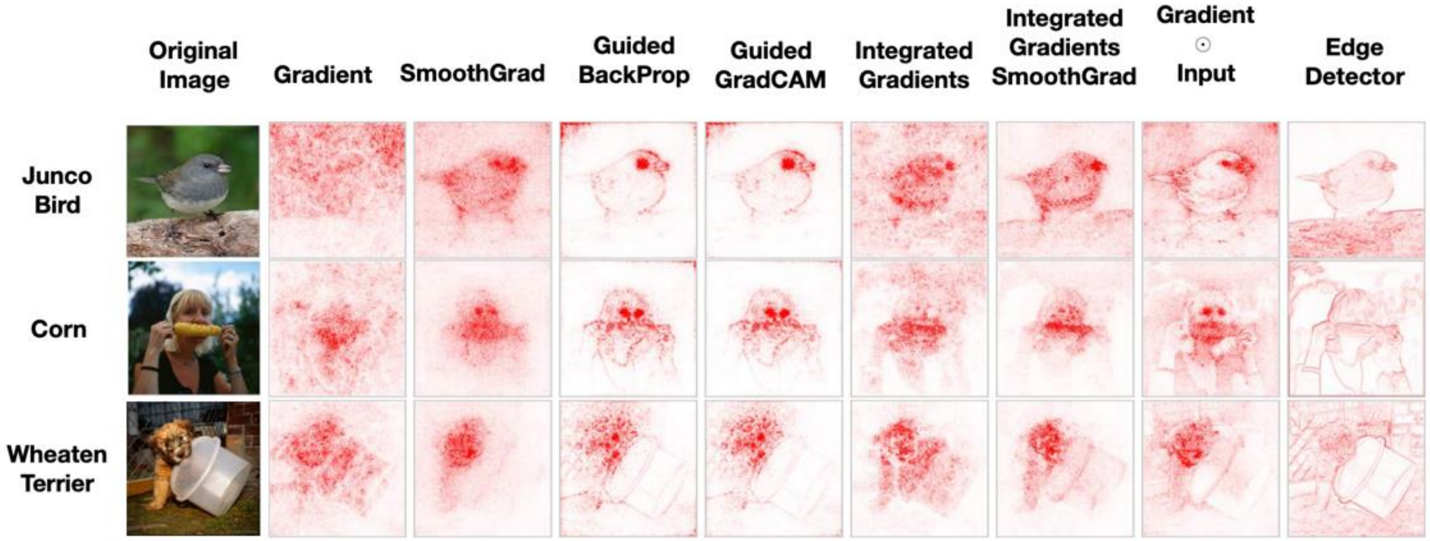
Case 3: Antidepressant prescribing

- Expert-generated ranking of ADs given patient scenarios = simulated ML model
- Systematically varied scenario, prediction accuracy, and explanation
- Any **explanation increased likelihood** to follow incorrect predictions
- Following incorrect predictions happened mostly with **feature-based explanations**



Jacobs et al., "How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection" 2021 *Translational Psychiatry*

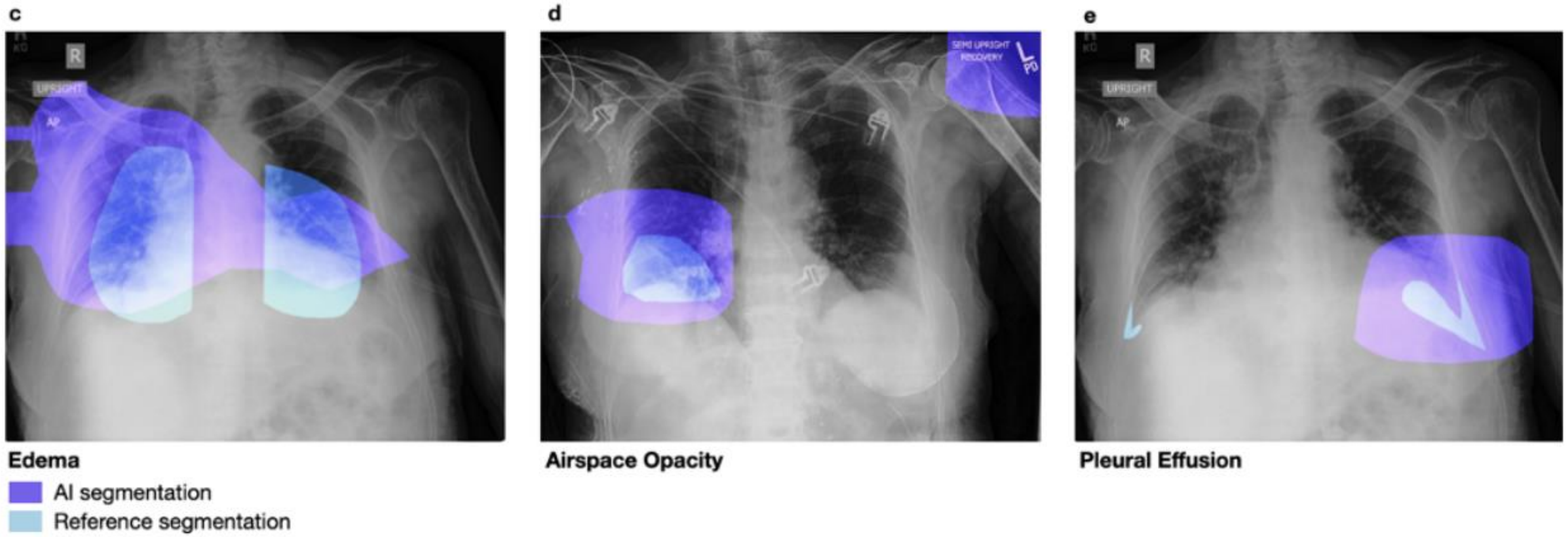
Are the explanations reliable?



- Explanations may not always provide accurate, relevant 'reasons' for their predictions[1]
- We assume that models are using information the same way that we do[2]

[1]Adebayo et al., 2018 "Sanity checks for saliency maps" 2018 *NeurIPS*; [2]Rudin "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead *Nat Med Intell*

CheXplanation



- Explanations highlight both relevant and non-relevant information
- Accuracy of explanations is correlated with model confidence
- … most reliable and most ‘correct’ in the clearest cases

[1]Saporta et al. “Deep learning saliency maps do not accurately highlight diagnostically relevant regions for medical image interpretation.” medRxiv. 2021 Jan 1.

THE LANCET
Digital Health

VIEWPOINT | VOLUME 3, ISSUE 11, E745-E750, NOVEMBER 01, 2021

PDF [899 KB]   Figures

The false hope of current approaches to explainable artificial intelligence in health care

Marzyeh Ghassemi, PhD • Luke Oakden-Rayner • Andrew L Beam, PhD

Open Access • Published: November, 2021 • DOI: https://doi.org/10.1016/S2589-7500(21)00208-9 •

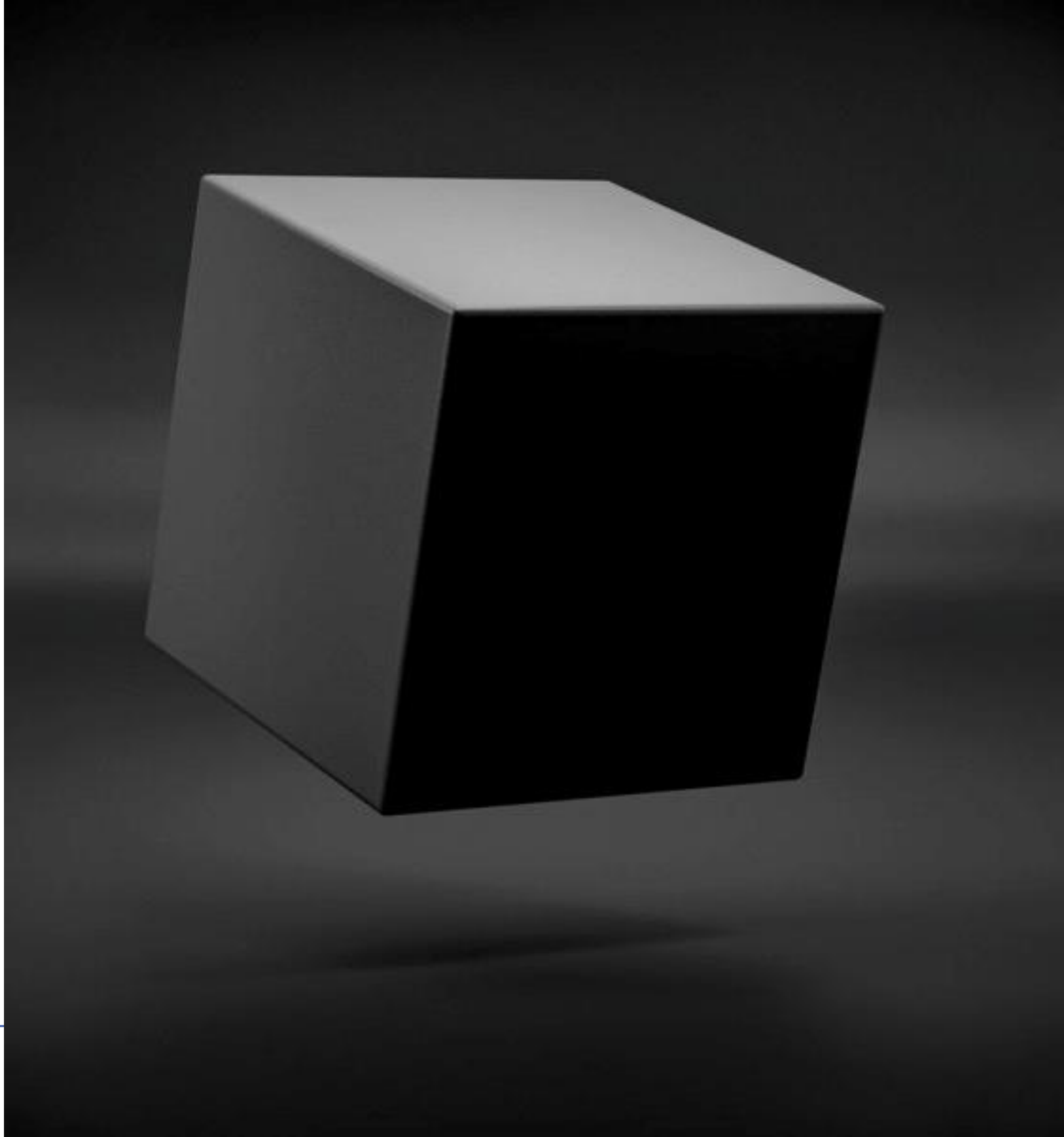
Check for updates

Summary

The black-box nature of current artificial intelligence (AI) has caused some to question whether AI must be explainable to be used in high-stakes scenarios such as medicine. It has been argued that explainable AI will engender trust with the health-care workforce, provide transparency into the AI decision making process, and potentially mitigate various kinds of bias. In this Viewpoint, we argue that this argument represents a false hope for explainable AI and that current explainability methods are unlikely to achieve these goals for patient-level decision support. We provide an overview of current explainability techniques and highlight how various failure cases can cause problems for decision making for individual patients. In the absence of suitable explainability methods, we advocate for rigorous internal and external validation of AI models as a more direct means of achieving the goals often associated with explainability, and we caution against having explainability be a requirement for clinically deployed models.

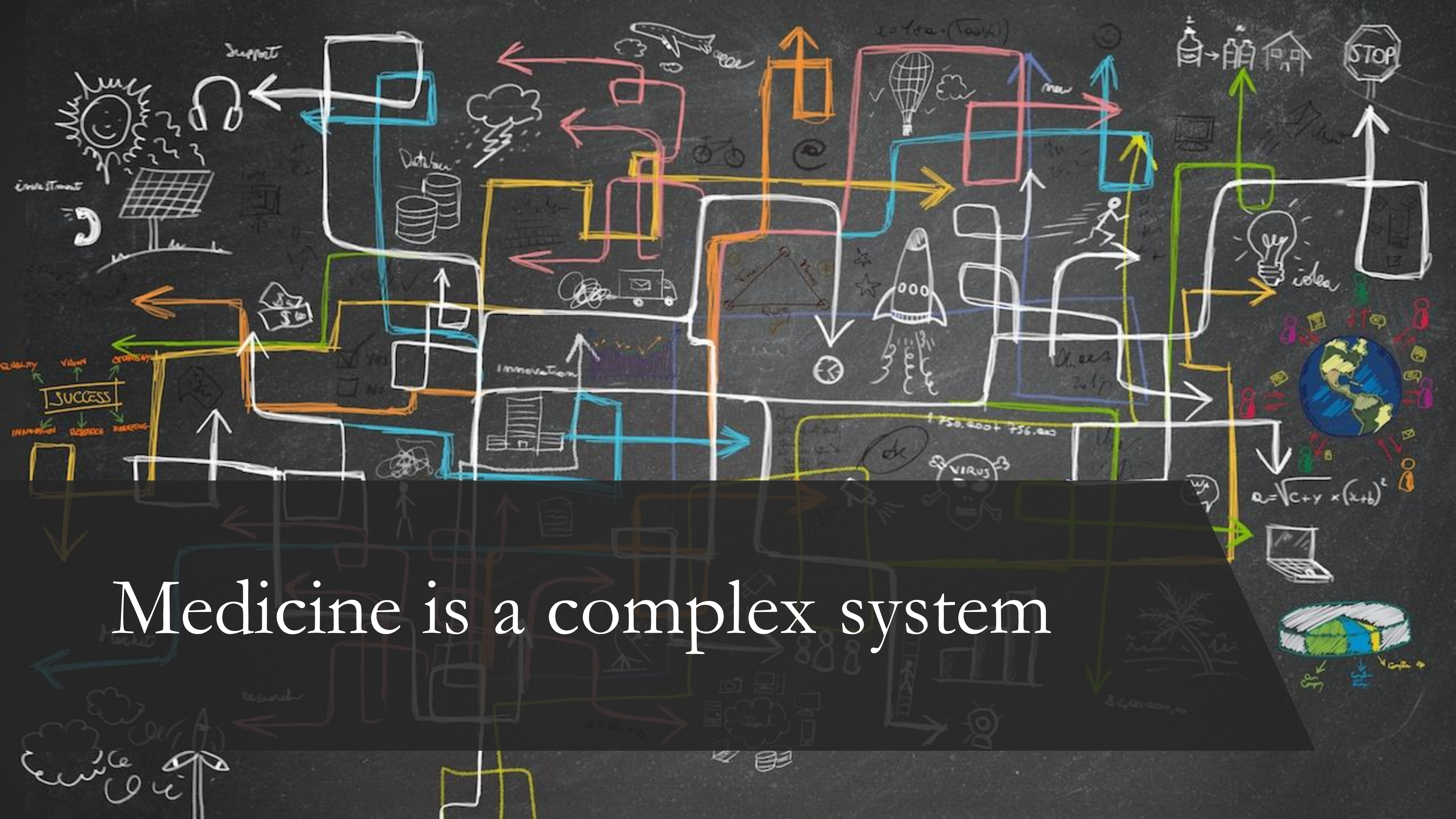
Is the black box really the problem?

- Clinicians are motivated to use the best available evidence to care for patients
- Evidence comes through clinical evaluations, not AUCs!
- If AI is believed to be superior as a form of knowledge, then it may seem reasonable to rely on its predictions
- But this is only part of the picture of how clinical decisions are made…



So how should we make AI-informed decisions?

Toward a humanistic vision of medicine augmented by AI



Medicine is a complex system

Multiple axes of knowledge

- ML formalizes one particular axis of knowledge in relation to a larger clinical decision that needs to be made[1]
  - E.g., antidepressant prescribing[2]
- "Clinical decisions are not made solely on the basis of the biological, physiological, and medical information supplied to the clinicians"[3] (Dr. Roxanne Kirsch)

[1]McCradden, 2021 *Transl Psych*; [2]Jacobs et al., 2021 *Transl Psych*; [3]McCradden & Kirsch, *in progress*

What is a ‘good’ decision with AI?

- Prospective qualitative study with clinicians (physicians, nurses, respiratory therapists) in intensive care, emergency medicine, and other acute care settings, and machine learning experts
- Case of ‘Siri’ simulated a typical handover in the ICU at SickKids
- Siri (4mos, 5kg):
  - Ventricular septal defect (Repaired); grade 2 subglottic stenosis and tracheomalacia; Trisomy 18
- Participants asked about their plan for Siri for the day wrt exubation
- Offered prediction of extubation readiness from simulated ML model

McCradden MD, Thai K, Assadi A, Tonekaboni S, Joshi S, Chevalier F, Zhang M, Goldenberg A. *In progress*

What is a ‘good’ decision with AI?

(1) Model card: documentation of general performance characteristics and use-case (Mitchell et al).

Model Facts

Model name: Extubation Prediction
Version: 1.0

Summary
This ML model uses EHR data collected from pediatric ICU encounters to predict real-time risk of extubating intubated patients using a Deep Neural Network. The model has been trained in house at SickKids trained on data retreived from 10/18-10/19

Mechanism
- **Outcome**........Real-time risk of extubation failure
- **Output**........0%-100% Risk of extubation failure
- **Patient population**........Age 0-5 admitted to SickKids PICU
- **Time of prediction**........Realtime
- **Input data source**........SickKids EPIC EHR
- **Input data type**........demographics, vitals, labs, interventions
- **Training site and time period**........SickKids EHR 10/19-05/20
- **Model type**........Deep Neural Network

Validation and performance
- **Retrospective**........20% held-out data from 10/18-10/19, AUROC: 0.90
- **Temporal**........5 months validation from 11/2019-04/2020, AUROC: 0.88
- **Prospective**........On-site AUROC(as of 15/05/20): 0.85
- **External**........FDA Approved (28/09/20)

Uses and directions
- **General Use:** The model is intended to be used as a decision support tool to assess the realtime risk of extubation. Clinical decision should be taken in conjunction with additional risk assessment, expert consulation, and additional patient information like X-rays, clinical notes, etc.
- **Examples of appropriate decisions of support:** Identifying potential risk of extubation
- **Before using this model:** Deployment outside of SickKids PICU should be subject to prospective studies on site for appropriate testing of generalizability and evaluated within the hospital clinical workflow
- **Effectiveness of evaluation:** Prospective evaluation over a month at SickKids PICU determined utility in assessing realtime extubation risk
- **Safety Evaluation:** TBD

Warnings
- **General Warnings:** Model is not to be used on general adult populations, ED or outside of SickKids. All risk estimates should be evaluated with other patient context. Model does not account automatically for past patient ICU stays in assessing risk. Model may not provide accurate risk estimates under rare pre-existing conditions.
- **Examples of inappropriate decisions of support:** Executing extubation without further patient assessment
- **Disccontinue use if:** Consistent contradictory performance is observed in relation to clinical assessment, changes in EHR system, or significant changes in pediatric population presenting to the ICU e.g. under outbreaks and other unprecedented circumstances

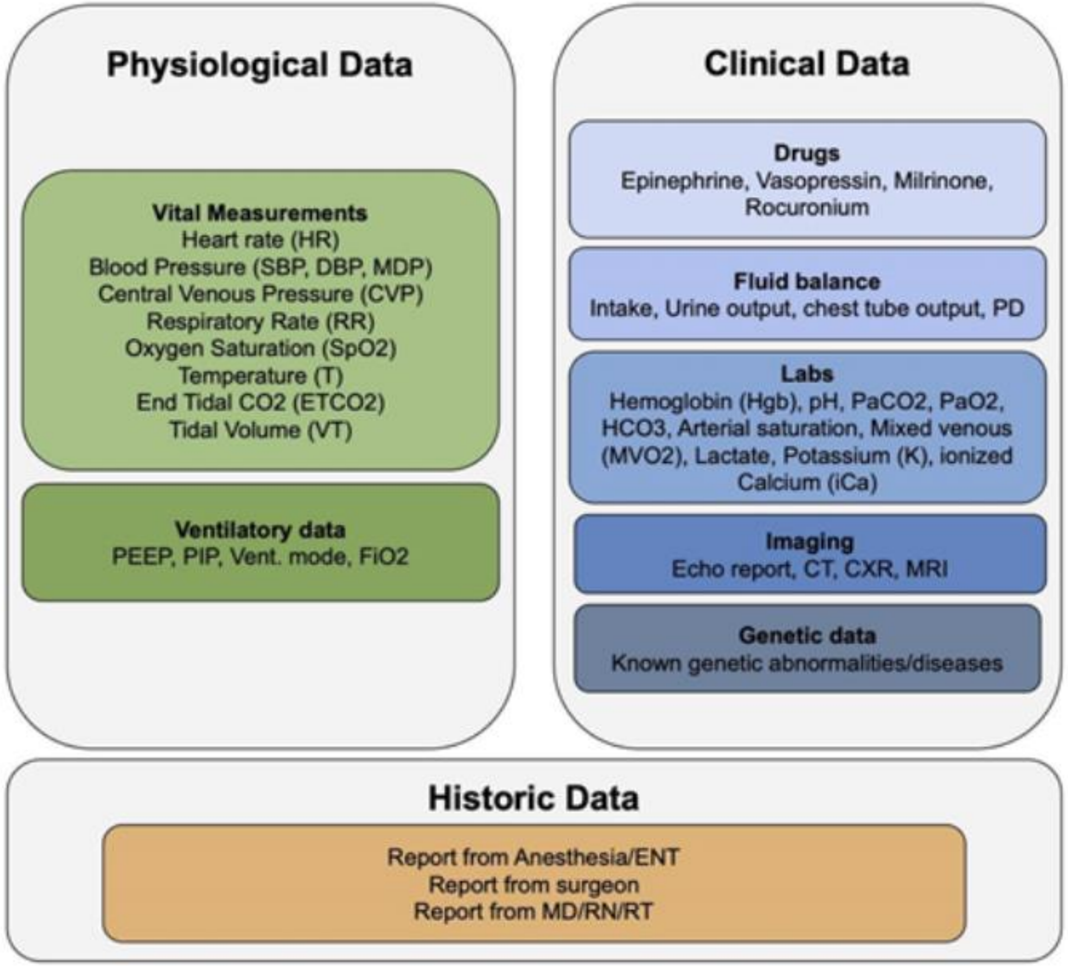
Other information
Publication:
1) Preprint - https://arxiv.org/abs/1905.05134.
Subgroup performance:
Prospective evaluation passess validation on major subgroups recognized by SickKids
Extubation procedure:

SickKids

McCradden MD, Thai K, Assadi A, Tonekaboni S, Joshi S, Chevalier F, Zhang M, Goldenberg A. *In progress*

What is a ‘good’ decision with AI?

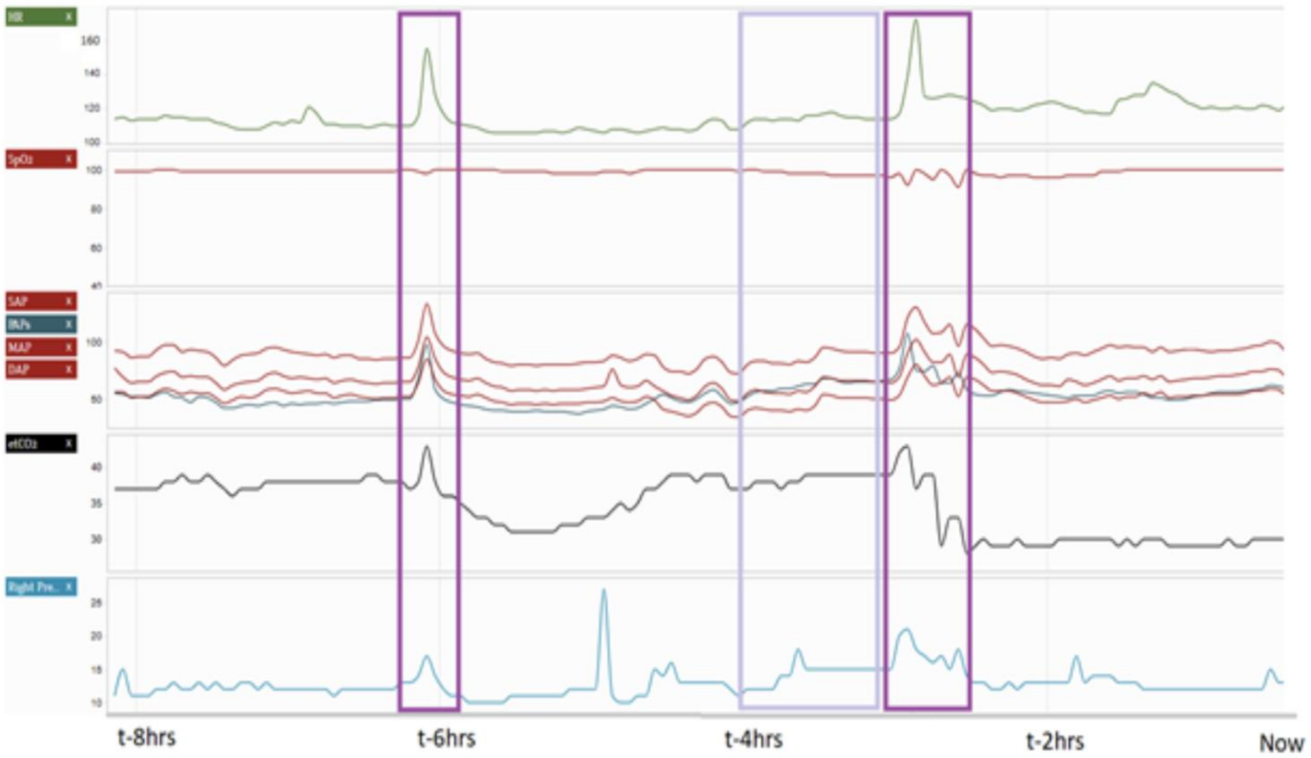
(2) Feature importance: Quantifying the influence of each input feature on the individual prediction (Lundberg et al).



McCradden MD, Thai K, Assadi A, Tonekaboni S, Joshi S, Chevalier F, Zhang M, Goldenberg A. *In progress*

What is a ‘good’ decision with AI?

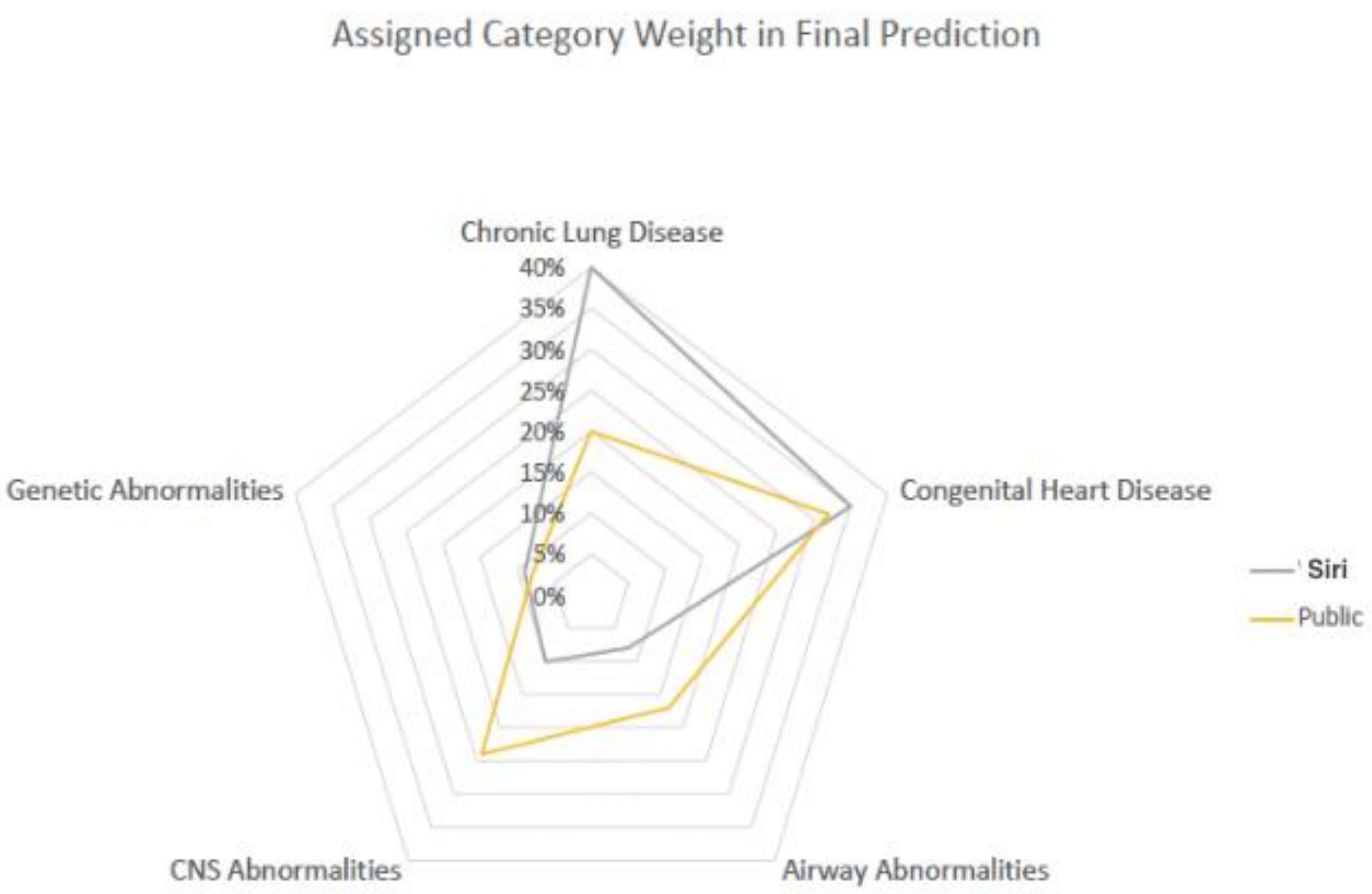
(3,4) Temporal explanation: Quantifying the influence of each input feature at different points in time on the individual prediction (Tonekaboni et al, Hardt et al)



McCradden MD, Thai K, Assadi A, Tonekaboni S, Joshi S, Chevalier F, Zhang M, Goldenberg A. *In progress*

What is a ‘good’ decision with AI?

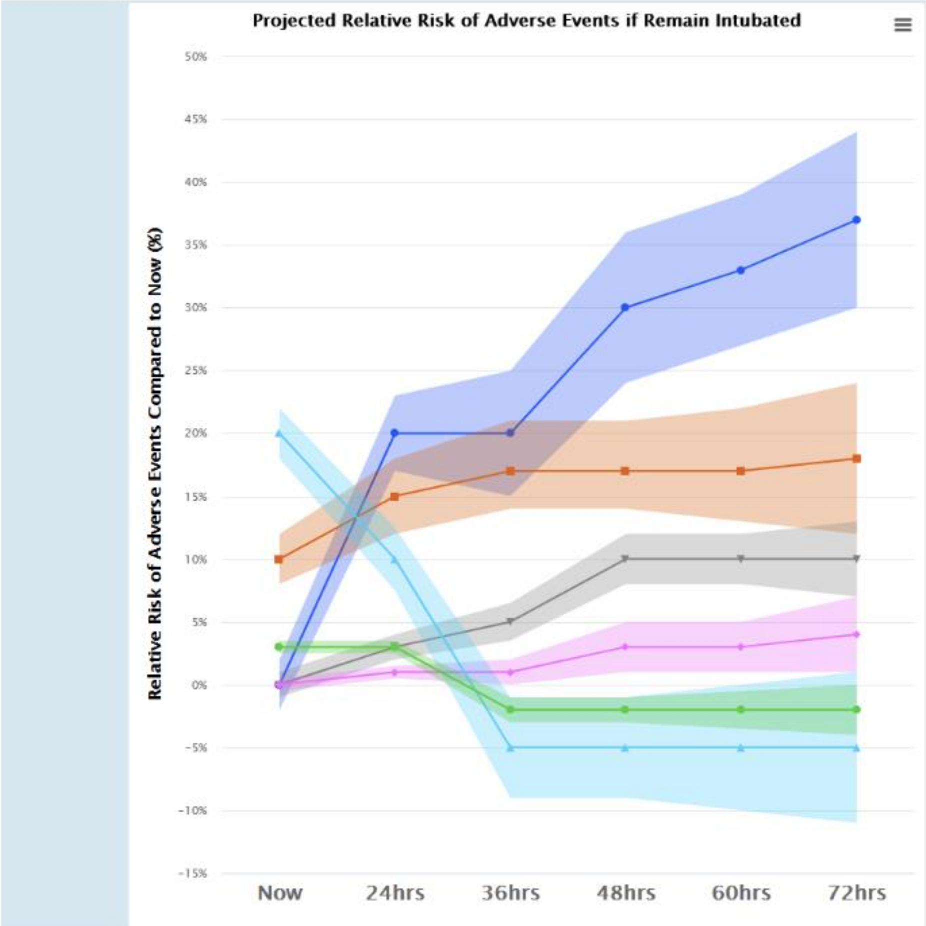
(5) Population-level explanation:
Contextualizing the individual model prediction in the training population



McCradden MD, Thai K, Assadi A, Tonekaboni S, Joshi S, Chevalier F, Zhang M, Goldenberg A. *In progress*

What is a ‘good’ decision with AI?

(6) Counterfactual/forecasting explanation: Estimating future outcomes depending on an action (Ates et al, Delaney et al)



	Now	24hrs	36hrs	48hrs	60hrs	72hrs
Sedative Use	0%	20%	20%	30%	33%	37%
Accidental Extubation	10%	15%	17%	17%	17%	18%
VAP	0%	3%	5%	10%	10%	10%
Deconditioning	0%	1%	1%	3%	3%	4%
Respiratory Failure	20%	10%	-5%	-5%	-5%	-5%
Cardiac Failure	3%	3%	-2%	-2%	-2%	-2%

McCradden MD, Thai K, Assadi A, Tonekaboni S, Joshi S, Chevalier F, Zhang M, Goldenberg A. *In progress*

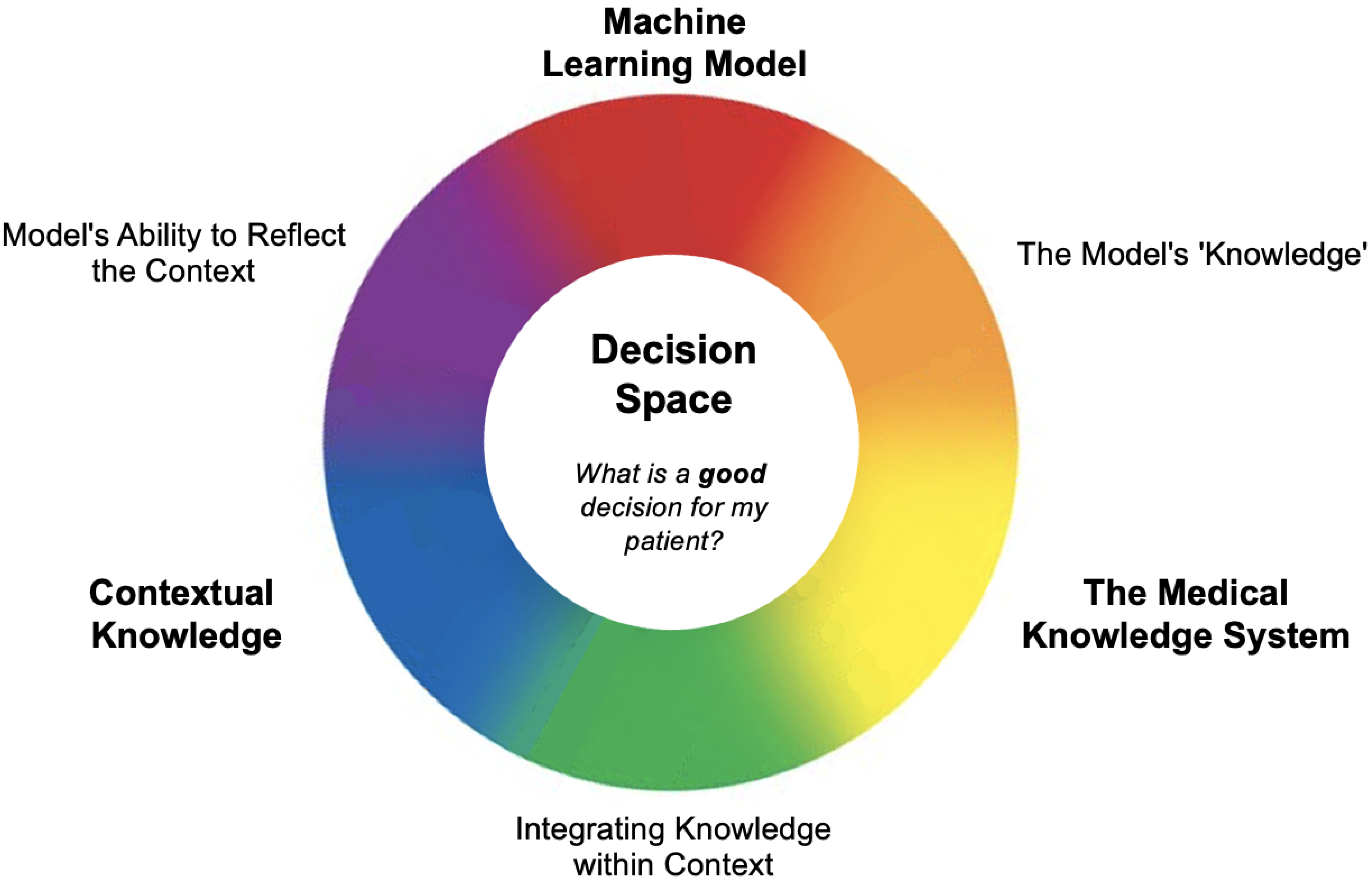
What is a ‘good’ decision with AI?

(7) High-level feature importance explanation

A success rate of 60% is predicted for this patient based on the following risk factors:

- T18
- Grade 3 airway with recent ENT intervention/involvement
- Baseline need for respiratory support
- Pulmonary edema on CXR
- Periodic episodes of vital sign instability (2SD deviation from expected values for age)

Figure 1: How clinicians can make a good decision using AI tools



This figure represents the theoretical model for making good decisions using AI. The outer circle represents the domains required by the clinician to reflect upon in order to make a responsible decision for an individual patient. A resolution is achieved through reflective equilibrium across all domains, using the goals of care as the guidepost.

Making morally good decisions

1. AI systems evaluated prospectively in a clinical environment
2. Information generated through this evaluation is aligned with the informational needs of the clinicians using the model
3. Clinical judgment may be calibrated using the evidence generated supporting the AI system
4. Particular attention must be paid to the model's performance on particular patient subgroups
5. Patient and family goals and values remain the guideposts – always acting to use medical knowledge to further the interests of patients

Key to the vision of using AI to make medicine more 'human'


It’s not just about ‘right’ or ‘wrong.’
It’s about morally good decisions in a highly complex system.

Thank you!

@mmccradden

melissa.mccradden@sickkids.ca

<u>Collaborators:</u>

Shalmali Joshi
James Anderson
Elizabeth Stephenson
Anna Goldenberg
Randi Zlotnik Shaul
Mjaye Mazwi
Roxanne Kirsch
Sana Tonekaboni
Farzad Khalvati
Karima Karmali
Wayne Lee
Minfan Zhang
Fanny Chevalier
Alex John London

SickKids

National Institutes of Health
Office of Data Science Strategy

Perspectives on AI and the NIH Data Ecosystem

SNMMI AI SUMMIT

March 21, 2022

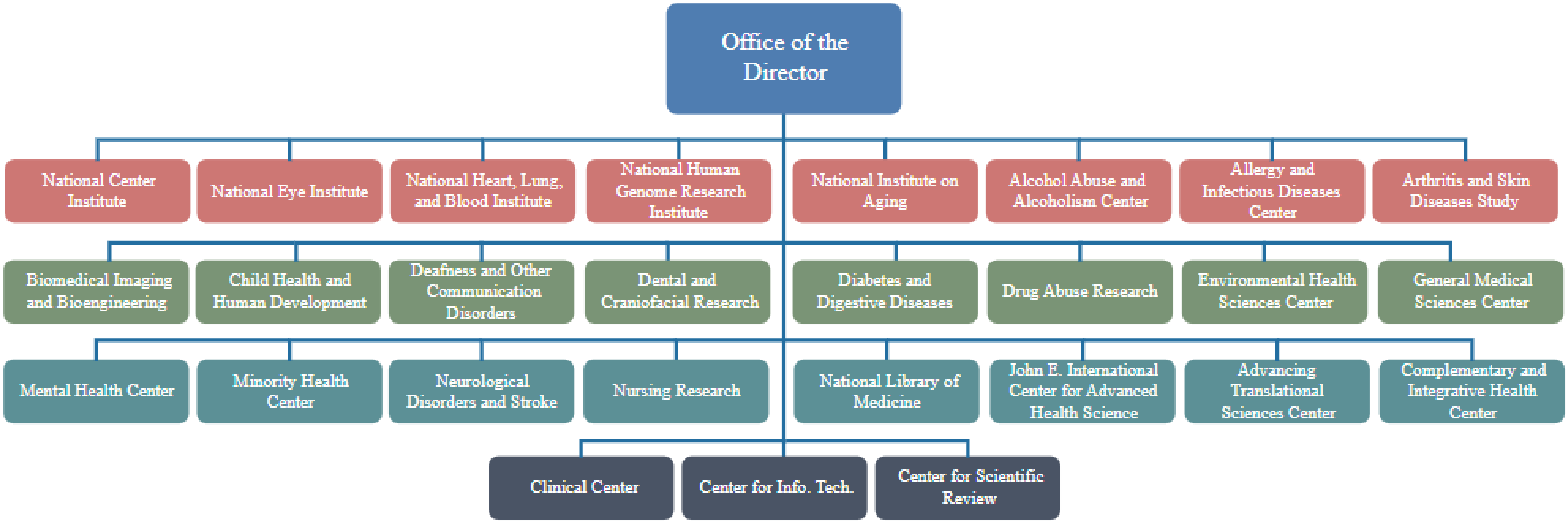
Laura Biven, Ph.D.
Lead, Integrated Infrastructure and Emerging Technologies
Office of Data Science Strategy
National Institutes of Health
Laura.Biven@nih.gov, https://datascience.nih.gov/

Outline

- NIH Office of Data Science Strategy
- NIH Data Ecosystem
  - Data
  - STRIDES
  - NCPI
- AI Activities
  - AIM-AHEAD
  - B2AI
  - ODSS

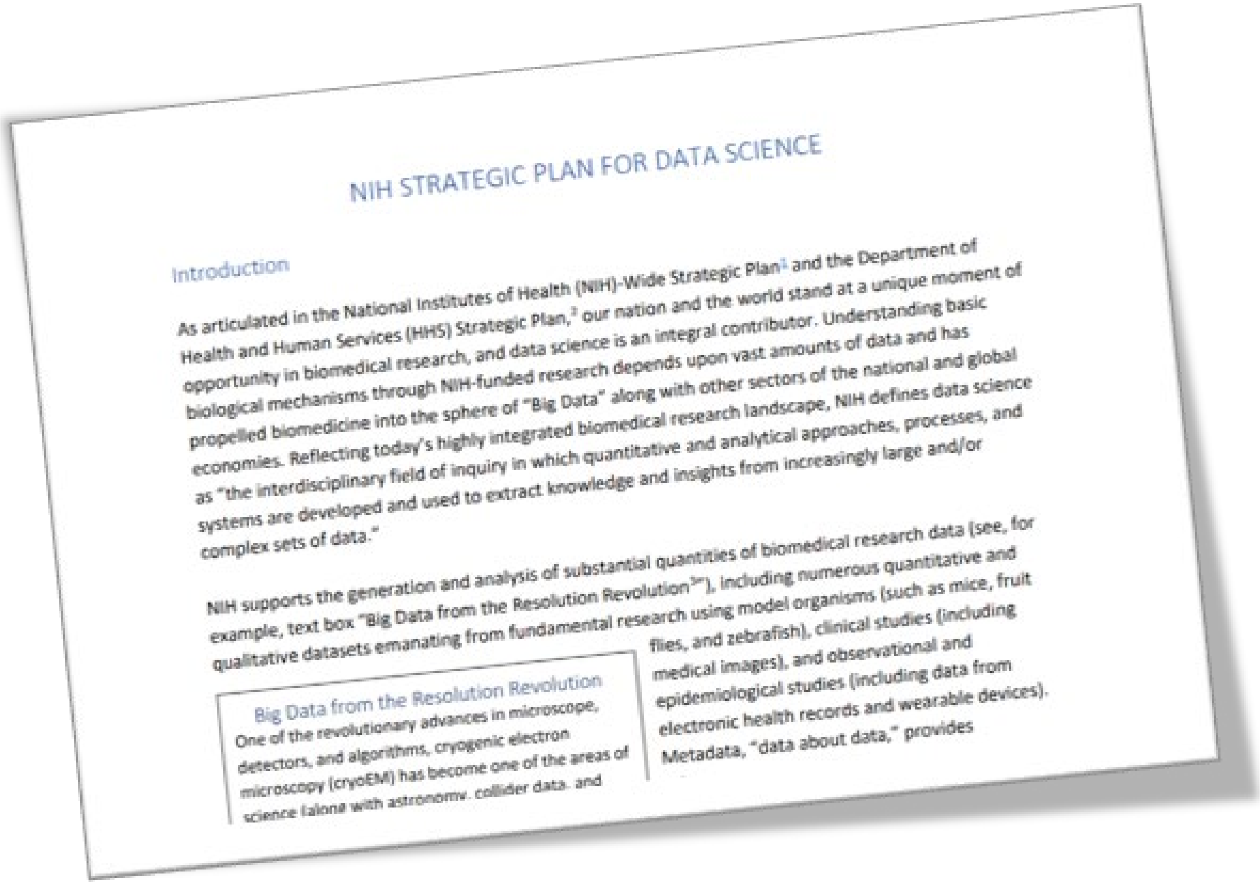
National Institutes of Health Institutes, Centers, and Offices

National Institute of Health (NIH) Org Chart



NIH Strategic Plan for Data Science

VISION:
A modernized, integrated, FAIR, biomedical data ecosystem

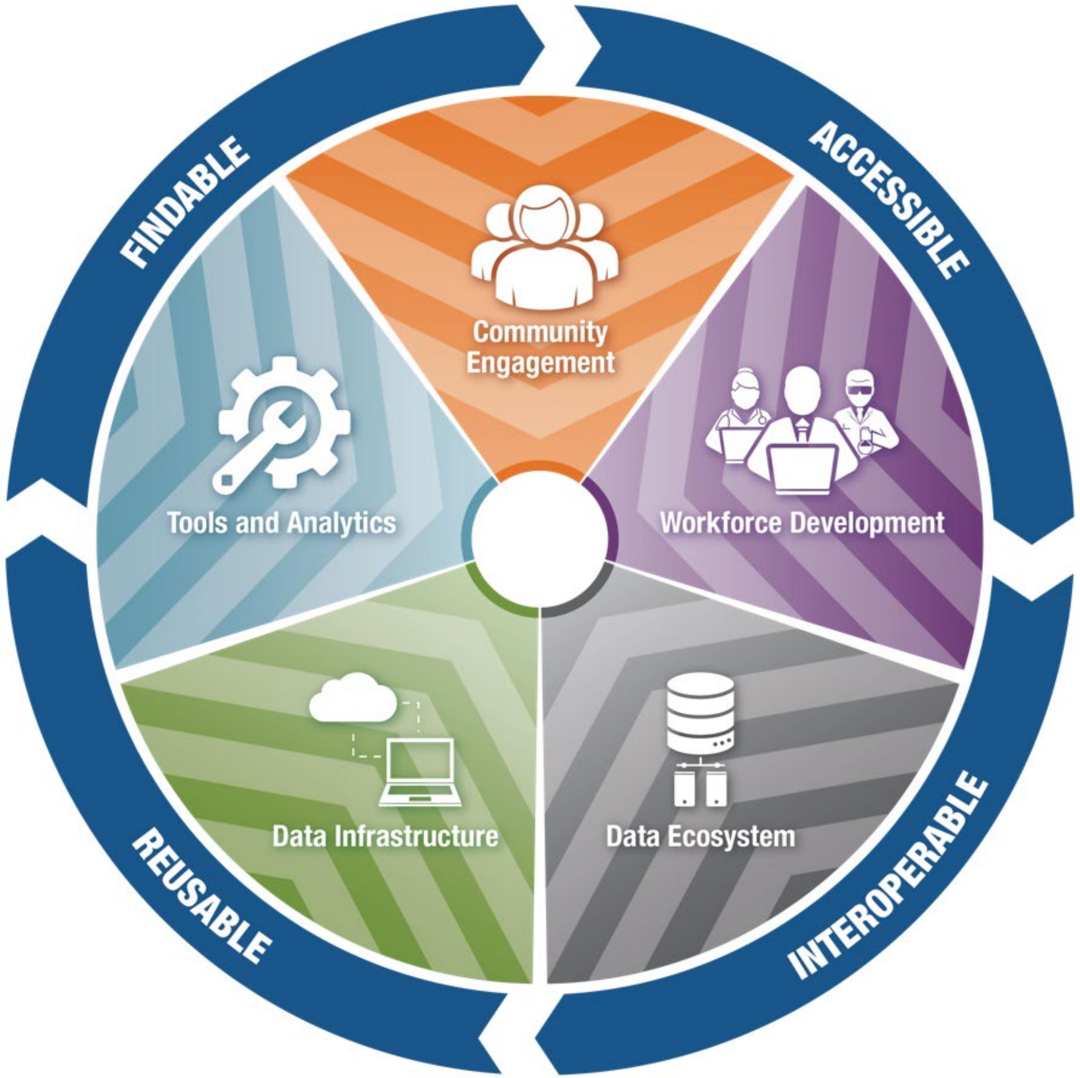

NIH STRATEGIC PLAN FOR DATA SCIENCE

Introduction

As articulated in the National Institutes of Health (NIH)-Wide Strategic Plan[1] and the Department of Health and Human Services (HHS) Strategic Plan,[2] our nation and the world stand at a unique moment of opportunity in biomedical research, and data science is an integral contributor. Understanding basic biological mechanisms through NIH-funded research depends upon vast amounts of data and has propelled biomedicine into the sphere of "Big Data" along with other sectors of the national and global economies. Reflecting today's highly integrated biomedical research landscape, NIH defines data science as "the interdisciplinary field of inquiry in which quantitative and analytical approaches, processes, and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data."

NIH supports the generation and analysis of substantial quantities of biomedical research data (see, for example, text box "Big Data from the Resolution Revolution"), including numerous quantitative and qualitative datasets emanating from fundamental research using model organisms (such as mice, fruit flies, and zebrafish), clinical studies (including medical images), and observational and epidemiological studies (including data from electronic health records and wearable devices). Metadata, "data about data," provides

Big Data from the Resolution Revolution
One of the revolutionary advances in microscope, detectors, and algorithms, cryogenic electron microscopy (cryoEM) has become one of the areas of science (along with astronomy, collider data, and



https://datascience.nih.gov/

Strategic Plan for Data Science: Goals and Objectives

Data Infrastructure	***Modernized Data Ecosystem***	***Data Management, Analytics, and Tools***	***Workforce Development***	***Stewardship and Sustainability***
Optimize data storage and security	Modernize data repository ecosystems	Support useful, generalizable, and accessible tools	Enhance the NIH data science workforce	Develop policies for a FAIR data ecosystem
Connect NIH data systems	Support storage and sharing of individual datasets	Broaden utility of, and access to, specialized tools	Expand the national research workforce	Enhance stewardship
	Better integrate clinical and observational data into biomedical data science	Improve discovery and cataloging resources	Engage a broader community	

NIH: Distributed Heterogeneous Repository Ecosystem

Domain/Data-specific
Open Access Data Sharing Repositories
as a first choice.
https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Datasets up to **2 gigabytes**

PubMed Central

Stores publication-related supplemental materials and datasets directly associated publications.

Datasets up to **20 gigabytes**

Generalist Repositories

(GREI Program)

Datasets associated with publications or otherwise and links to PubMed.

High priority datasets, **petabyte-scale**

Cloud Partners

(STRIDES, RAS, ICOs, etc.)

Store and manage large scale, high priority NIH datasets.

Consistent with Desirable Characteristics: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html

Positioning Repositories for Data Sharing

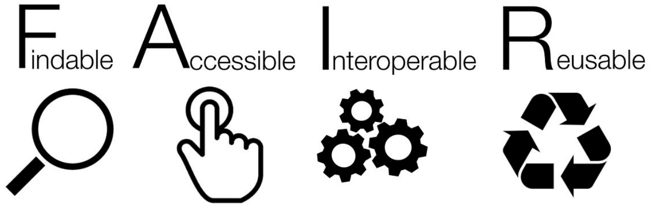
FY21: NOT-OD-21-089

FY22: NOT-OD-22-069

Support for existing data repositories to align with FAIR and TRUST principles and evaluate usage, utility, and impact

ODSS provided funding for existing repositories of all sizes, and at different stages of establishment to:

- Increase "FAIR"-ness and "TRUST"-worthiness
- Improve their usage, utility, and impact throughout the data resource lifecycle.



Data resources are key enablers of modern biomedical research. Awards promote data sharing by lowering barriers and reducing or eliminating silos. These shifts allows for the discovery and use of data, enabling better secondary use of data. ODSS promotes the implementation of best practices, increases reproducibility of research, and optimizes efficiency of operations and costs for data resources

Optimized Funding for NIH Data Repositories and Knowledgebases

- Data resources are important research tools
- Historically funded through research grants
- Funding mechanism should be optimal for type of resource
- **End goal:** researcher confident in data and information integrity

- **Solution:** New Funding Announcement for data repositories and knowledgebases
- Resource plan requirement

Scientific Impact

Community Engagement

Quality of Data and Services and Efficiency of Operations

Governance

PAR-20-089 and PAR-20-097

NEW: The Generalist Repository Ecosystem Initiative

Solicit applications from generalist repositories working together to:

- Implement consistent capabilities (NOT-OD-21-016)
- Create better access to & discovery of NIH funded data
- Conduct outreach & train on FAIR data practices
- Engage the research community

Expected Outcomes

- Make data sharing easier
- Improve discoverability
- Increase reproducibility of research
- Encourage secondary use of data

The Dataverse Project · DRYAD · figshare · MENDELEY DATA · OSF · Vivli CENTER FOR GLOBAL CLINICAL RESEARCH DATA

Enhance the biomedical data-science research workforce through improved programs and novel partnerships.

STRIDES Initiative (*The Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability*) provides:

- State-of-the-art data storage and computational capabilities
- Training and education for researchers
- Innovative technologies such as artificial intelligence and machine learning
- Professional engineering and technical support

Partnerships with: Microsoft Azure Google Cloud aws

The STRIDES Initiative aims to help NIH and its institutions accelerate biomedical research by reducing barriers in utilizing commercial cloud services. This initiative aims to harness the power of the cloud to accelerate biomedical discovery. NIH and NIH-funded researchers can take advantage of STRIDES benefits.

Enroll Now

Gain access to

- Discounts on partner services
- Professional services consultations
- Access to training
- Potential collaborative engagements

>163
Petabytes of Data

201M
Compute Hours

>693
NIH & NIH-funded Research Programs/ Projects

$28M
Cost Savings

>4081
People Trained

https://datascience.nih.gov/strides

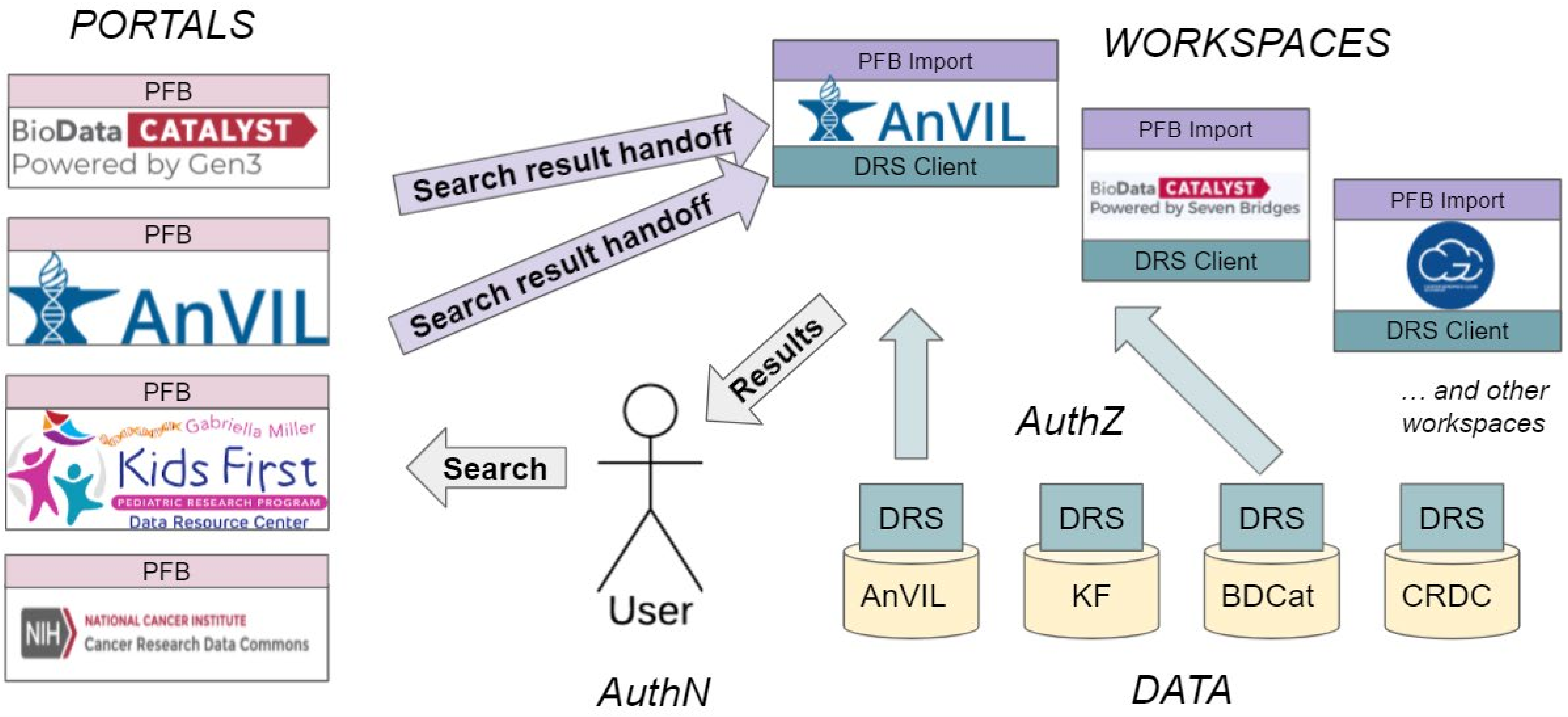
What is NCPI?

The NIH Cloud Platform Interoperability (NCPI) effort aims to establish and implement guidelines and technical standards to empower end-user analyses across participating NIH cloud platforms, to facilitate the realization of a trans-NIH, federated data ecosystem.

Established in late 2019 as a coalition of independently funded NIH IC cloud-based data platforms, with additional support from ODSS

https://anvilproject.org/ncpi

Diverse users can co-analyze data to drive science



Administrative Coordinating Center for the NIH Cloud Platform Interoperability (NCPI) Program

NCPI

- High-value NIH datasets are stored in multiple repositories hosted by individual institutes and centers
- **NCPI will enable a federated data eco-system to facilitate cross-platform data analysis**

OTA-22-004 NCPI Administrative Coordinating Center

NIH invites applications to provide technical, administrative, and coordination support for the NIH Cloud Platform Interoperability program (NCPI).

- LOIs due: April 15, 2022
- Full Proposals due: June 1, 2022

NIH ODSS SEARCH WORKSHOP

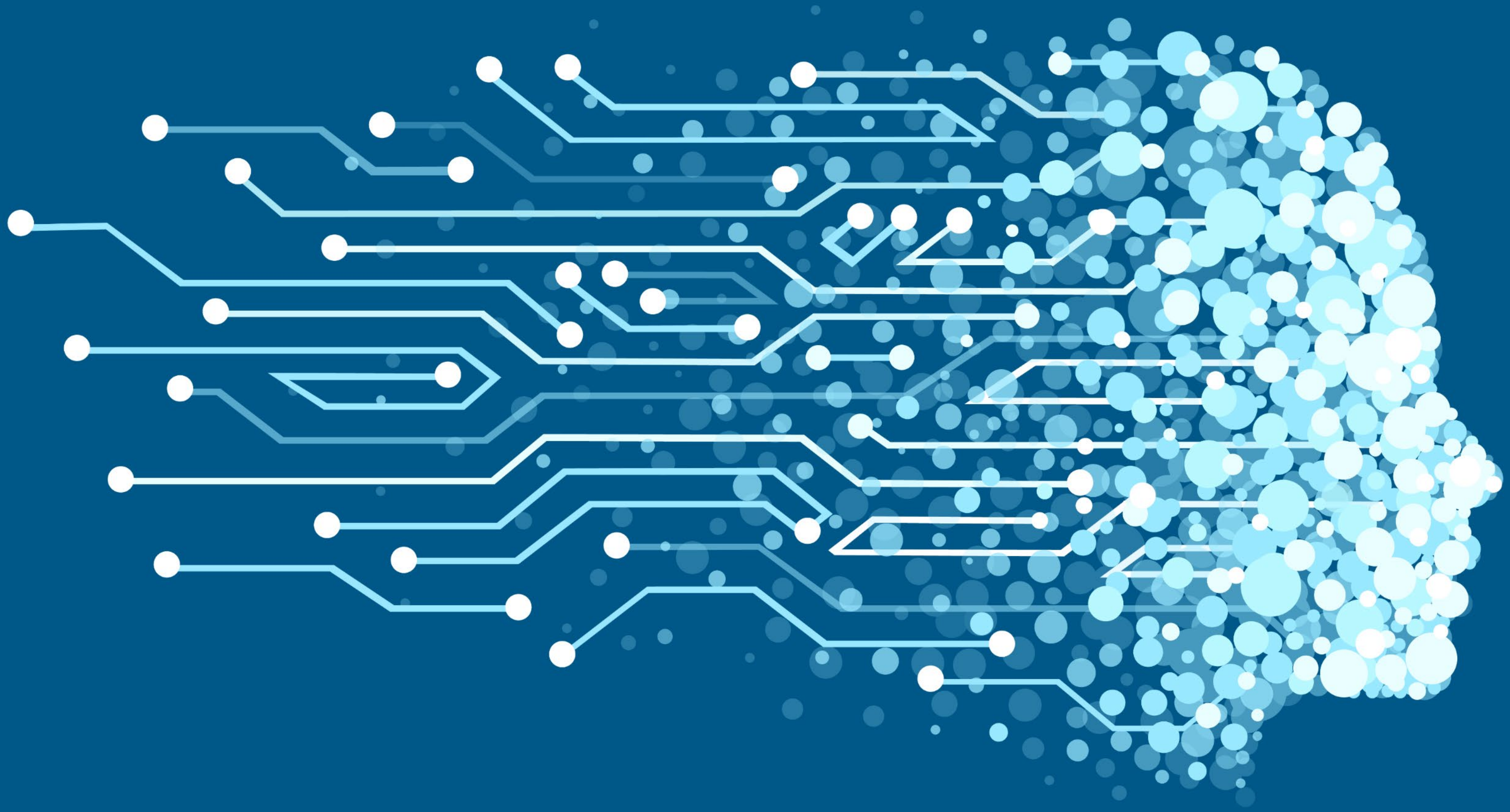
NIH National Institutes of Health
Office of Data Science Strategy

January 19-20, 2022

The Workshop explored current capabilities, gaps and opportunities for global data search across the data ecosystem with these main themes:

- Using search to build cohorts: finding data across different platforms/repositories using patient attributes in order to create a cohort of patients for clinical analysis
- Using search to find relevant data & repositories: finding data & repositories in order to access and analyze the data further, including its use for creating computational models.
- Using search for (complex) information retrieval: answering specific questions without the additional burden of data download or analysis

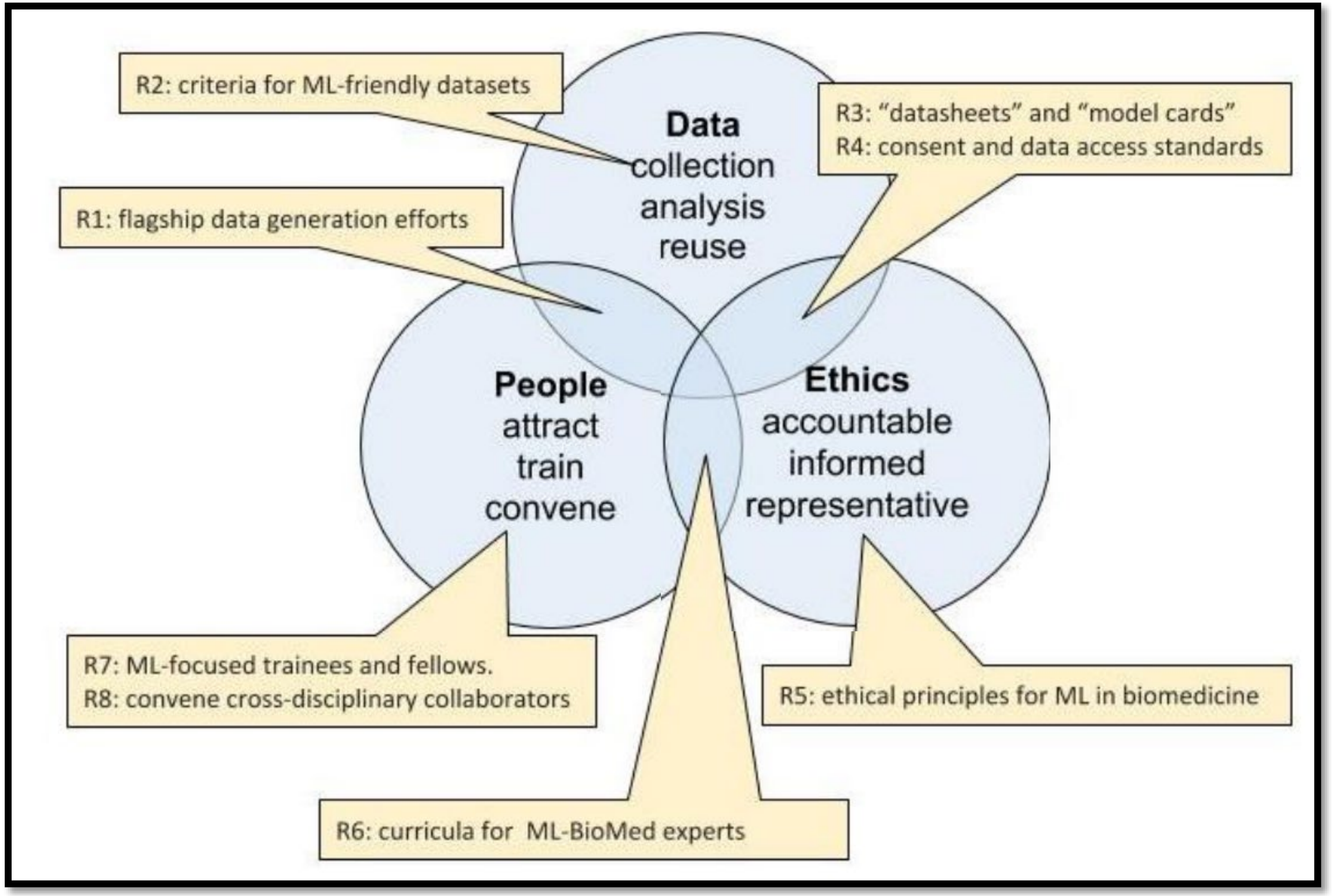
Link



AI

Biomedical AI: Visions for an **ETHICAL** Future

NIH ACD AI Working Group Recommendations:



- Outlined opportunities to fuse AI/ML with exponential increase in biomedical data
- Ethics was identified as equally important to Data and People, reflecting the primary importance of infusing ethical thinking into AI/ML use in biomedical research

https://www.acd.od.nih.gov/documents/presentations/12132019AI_FinalReport.pdf

Ethical AI/ML: A hot topic across federal agencies

OFFICE OF SCIENCE AND TECHNOLOGY POLICY

In the process of developing an **A.I. Bill of Rights**

NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE

OVERSEEING AND IMPLEMENTING THE UNITED STATES NATIONAL AI STRATEGY

Strategic Pillars: Innovation; Advancing Trustworthy AI; Education and Training; Infrastructure; Applications; International Cooperation

THE NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

"Americans have not yet grappled with just how profoundly the artificial intelligence (AI) revolution will impact our economy, national security, and welfare.…**The AI competition is also a values competition.**"

U.S. Department of Health and Human Services

Artificial Intelligence (AI) Strategy

Partnering with academia, industry and government, HHS will leverage AI to solve previously unsolvable problems by ***continuing to lead advances in the health and wellbeing of the American people***, *responding to the use of AI* across the health and human services ecosystem, and *scaling trustworthy AI adoption* across the Department.

Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD)

AIM AHEAD

Partnerships

Research

Infrastructure

Training

Goals:

- to enhance the **participation** and **representation** of researchers and communities currently underrepresented in the development of artificial intelligence and machine learning (AI/ML) models
- to address health disparities and inequities using AI/ML
- to improve the capabilities of this emerging technology, beginning with the use of electronic health record (EHR) and extending to other diverse data

https://aim-ahead.net/
https://datascience.nih.gov/artificial-intelligence/aim-ahead

Community Input Shaped the Initial Phase


Request for Information (RFI): Inviting Input to Broaden the Benefits of AI/ML Technologies to Reduce Health Disparities and Inequities and Enhance the Diversity of the AI/ML Workforce

Notice Number:
NOT-OD-21-147

Key Dates

Release Date:

Response Date:

Related Announcem

None

Issued by

Office of The Director, National Insti

Purpose

Through this Request for Information
Intelligence/Machine Learning (AI/ML
disparities and advance health equity
new data systems for underrepresen

Background

Artificial Intelligence/Machine Learning (AI/ML) Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD)

June 25, 2021

Event Details | Agenda

Registration Closed

Registration ended June 23, 2021

Meeting Objectives

This virtual forum aims to bring together stakeholders across academia, federal agencies, the data science/technology industry, and health care systems and centers who are interested in leveraging artificial intelligence and machine learning (AI/ML) for research with a primary focus on mitigating health disparities. The purpose of

Location

Webinar

The web link required to join the webinar

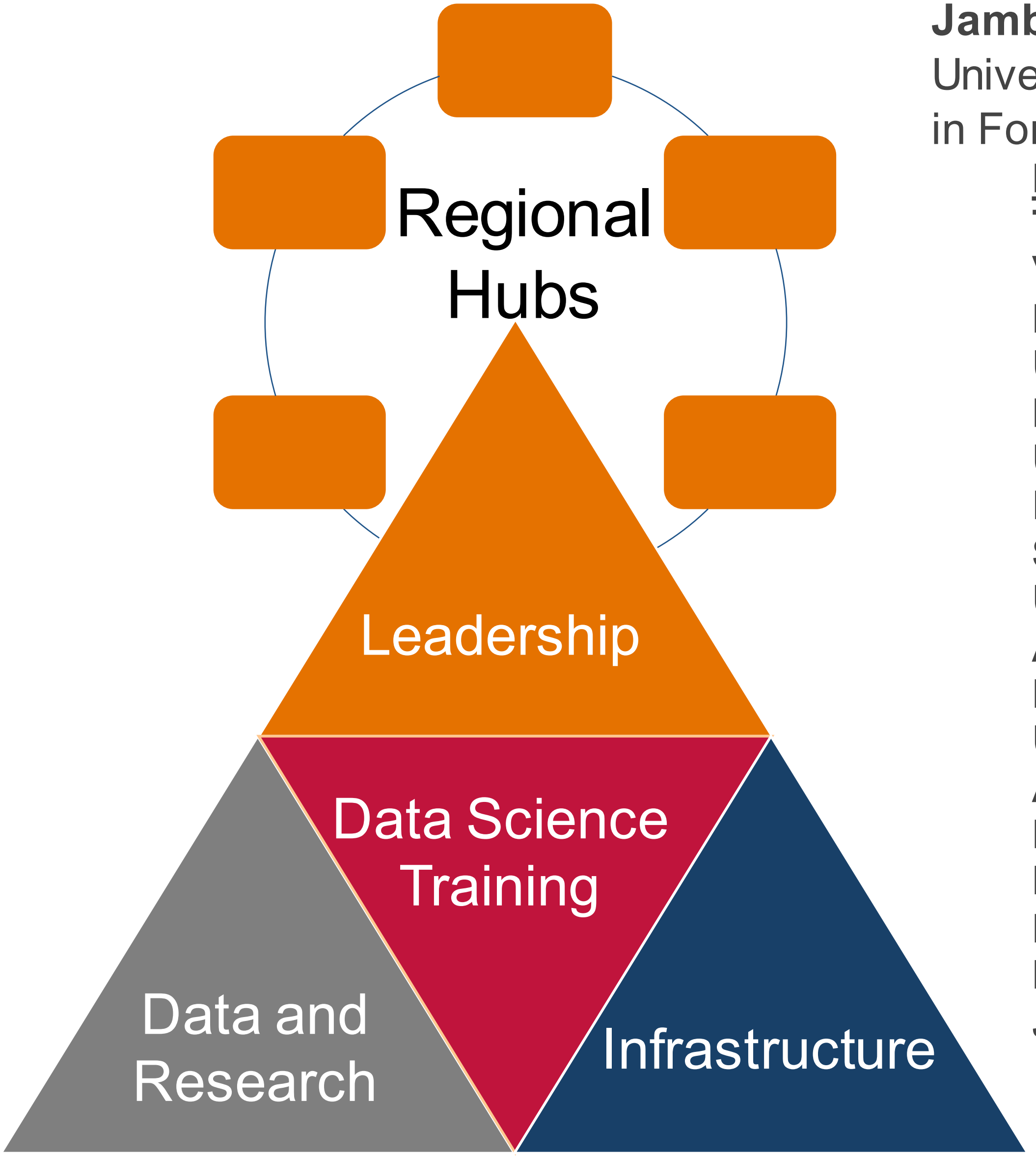
There is a wide variety of interests, needs, and resources across communities.

- AIM-AHEAD will develop a consortium of organizations and institutions that
  - wish to develop capabilities in AI/ML
  - wish to build a more inclusive basis for AI/ML
  - have a core mission to serve health disparity populations.

➔Begin with a two-year planning, assessment, and capacity building phase

➔Establish a Coordinating Center with the essential expertise in AI/ML and health disparities research, data science training, and data and computing infrastructure

The AIM-AHEAD Coordinating Center



Leadership Core

Jamboor K. Vishwanatha, Ph.D.
University of North Texas Health Science Center in Fort Worth

Regional Hubs

Toufeeq Ahmed, Ph.D.
Vanderbilt University Medical Center
Bettina Beech, Dr.P.H.
University of Houston
Harlan P. Jones, Ph.D.
University of North Texas Health Science Center in Fort Worth
Spero Manson, Ph.D.
University of Colorado-Anschutz Medical Center in Aurora
Keith Norris, M.D., Ph.D.
University of California, Los Angeles
Anil Shanker, Ph.D.
Meharry Medical College in Nashville, Tennessee
Herman Taylor, M.D.
Morehouse School of Medicine in Atlanta, Georgia
Roland J. Thorpe, Jr., Ph.D.
Johns Hopkins University in Baltimore, Maryland

Data Science Training Core

Legand L. Burge, Ph.D.
Howard University in Washington, D.C.

Infrastructure Core

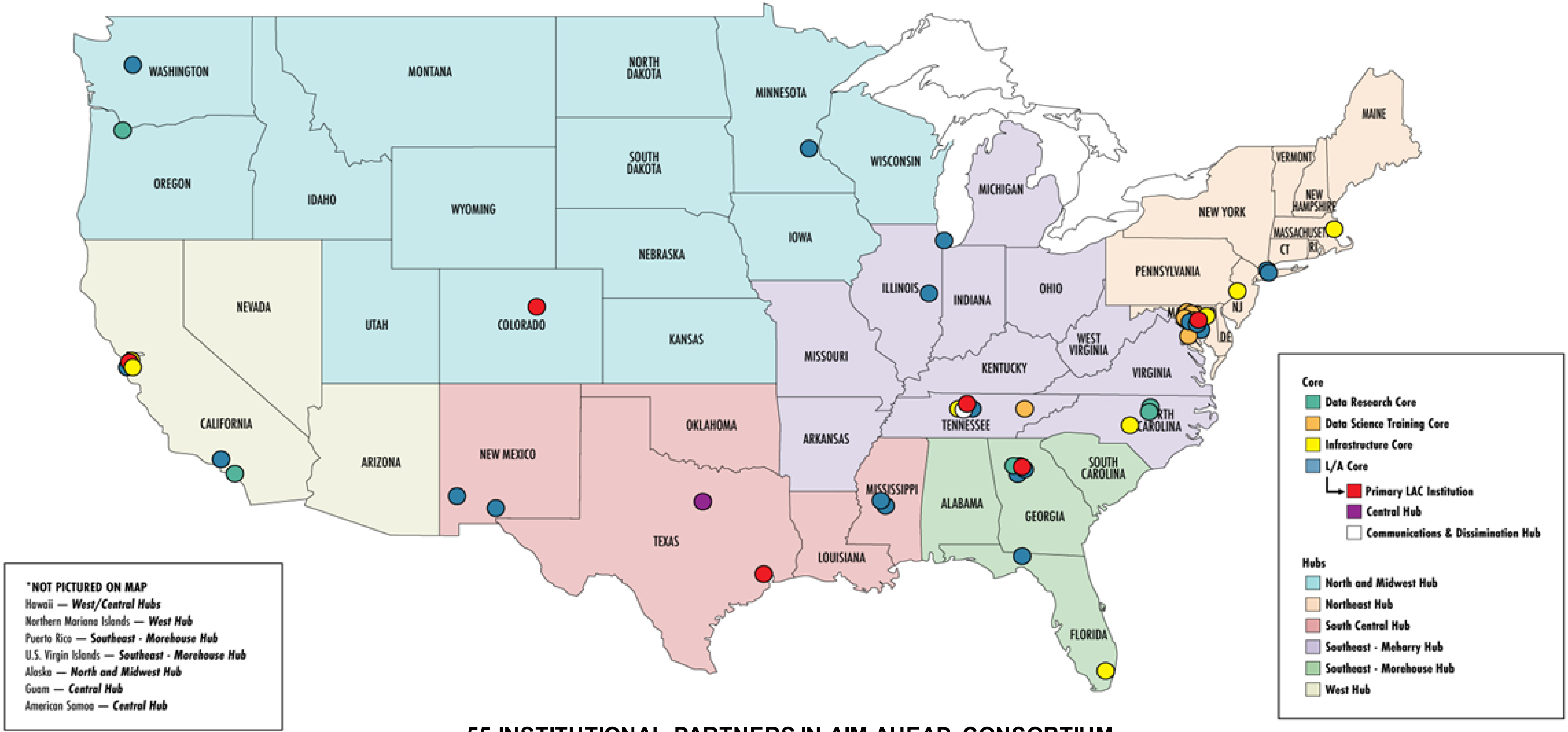
Alex J. Carlisle, Ph.D.
National Alliance Against Disparities in Patient Health in Woodbridge, Virginia
Paul Avillach, M.D., Ph.D.
Harvard Medical School in Boston, Massachusetts
Bradley A. Malin, Ph.D.
Vanderbilt University Medical Center in Nashville, Tennessee

Data and Research Core

Jon Puro, M.P.A.
OCHIN in Portland, Oregon

https://aim-ahead.net/

AIM-AHEAD Partnership Map



55 INSTITUTIONAL PARTNERS IN AIM-AHEAD CONSORTIUM

AIM-AHEAD

Artificial **I**ntelligence/**M**achine Learning Consortium
to **A**dvance **H**ealth **E**quity **a**nd Researcher **D**iversity

"I never anticipated the appetite for this initiative in [the American Indian, Alaskan Native, and Hispanic] communities… There is a thirst for this."

-- Spero Manson (Pembina Chippewa), Distinguished Professor of Public Health and Psychiatry, Director for the Centers for American Indian and Alaska Native Health, Associate Dean of Research at the Colorado School of Public Health at the University of Colorado Denver's Anschutz Medical Center

"…we need to build on programs like the new NIH AIM-AHEAD (or at least ensure their funding continues), to not only make sure diversity is covered in biomedical data sets, but diversity is promoted and enhanced among the data scientists themselves."

-- Atul Butte, MD, PhD, Priscilla Chan, and Mark Zuckerberg Distinguished Professor, University of California, San Francisco, Director, Bakar Computational Health Sciences Institute and Chief Data Scientist, University of California Health

Bridge2AI

- Use biomedical and behavioral research grand challenges to generate **flagship data sets**
- Emphasize **ethical** best practices
- **Prepare** AI/ML-friendly data
- Promote **diverse teams**

Standardize Data Attributes

Develop Automated Tools

New Datasets

Create cross-training materials for Workforce Development

Disseminate Products & Best Practices

https://commonfund.nih.gov/bridge2ai

Instilling a culture of ethical inquiry

Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25, 44–56 (2019). https://doi.org/10.1038/s41591-018-0300-7

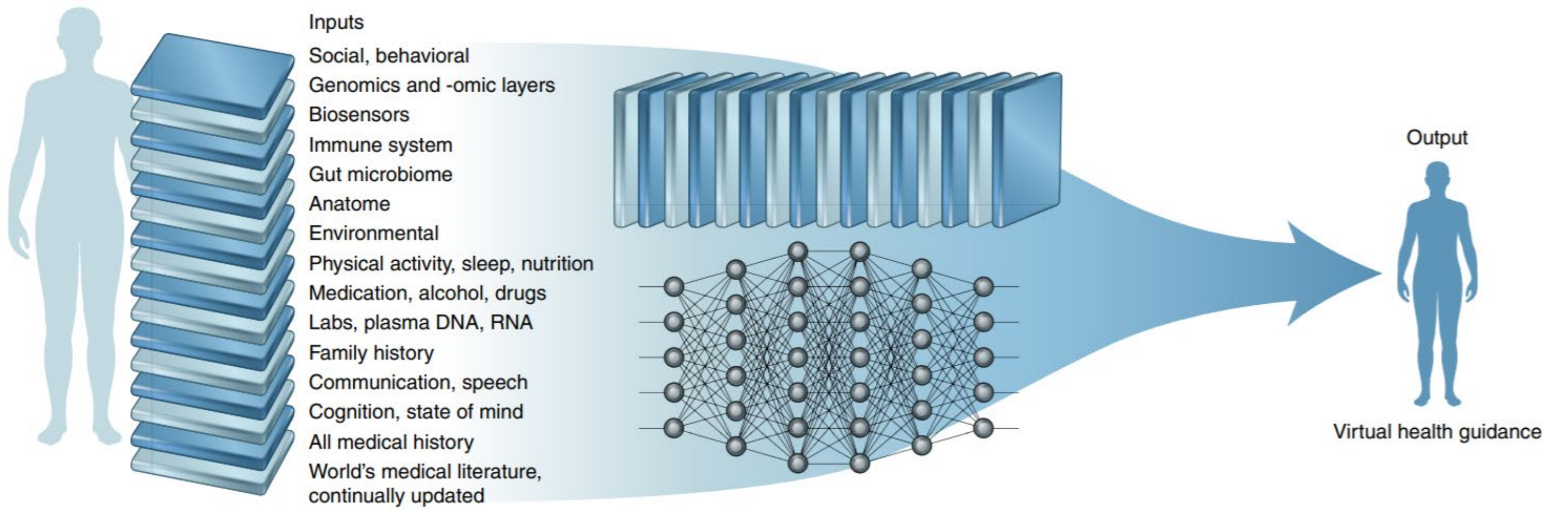
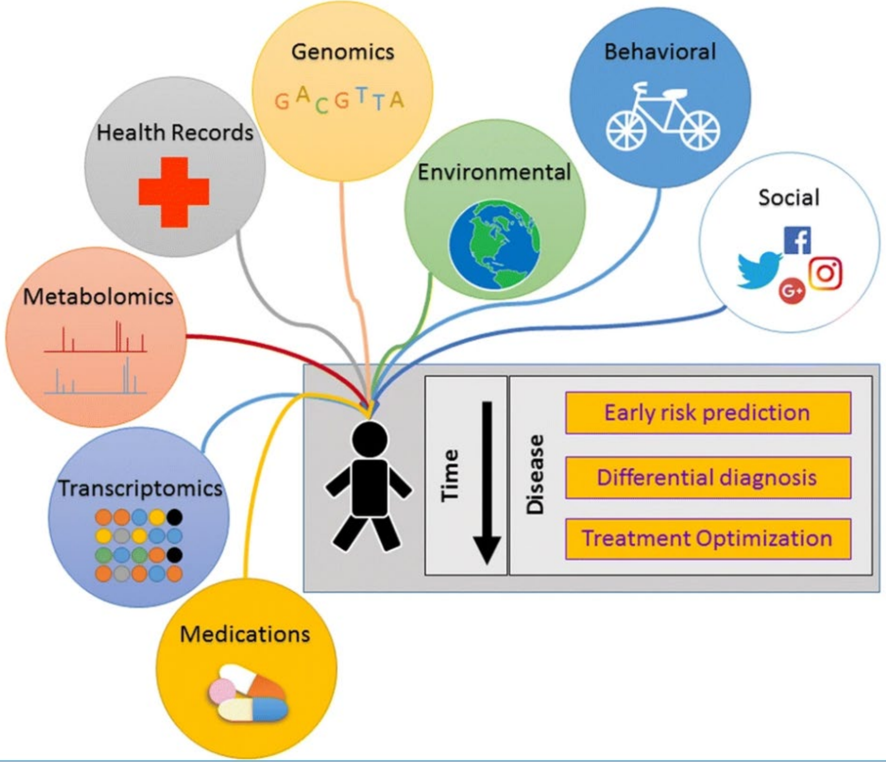


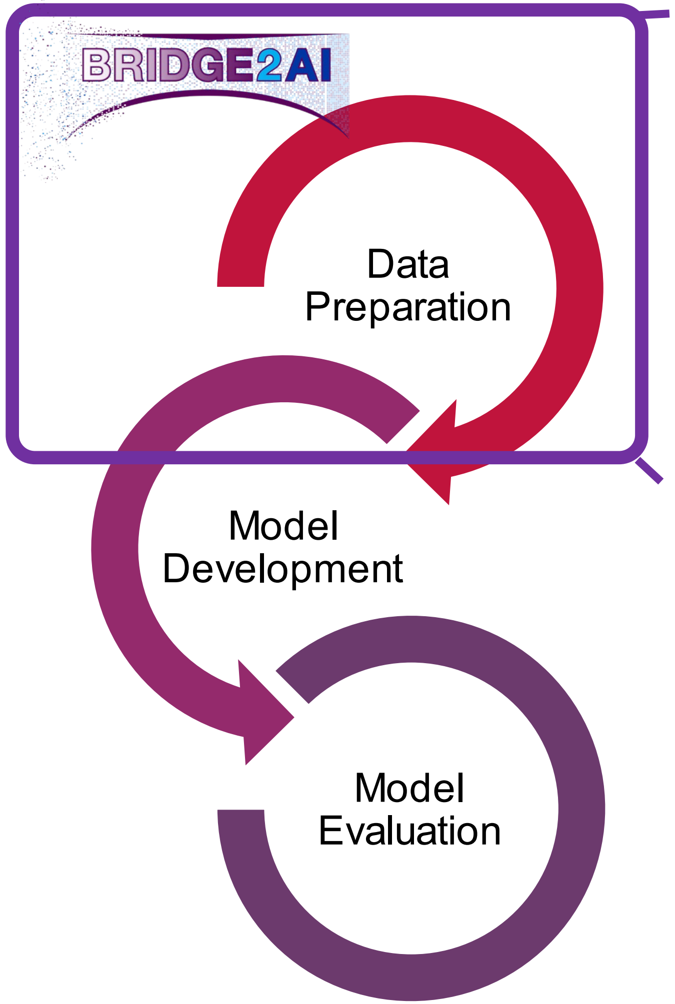
Fig. 3 | The virtual medical coach model with multi-modal data inputs and algorithms to provide individualized guidance. A virtual medical coach that uses comprehensive input from an individual that is deep learned to provide recommendations for preserving the person's health. Credit: Debbie Maizels/ Springer Nature

To integrate all types of ethically sourced biomedical and behavioral data to predict health outcomes

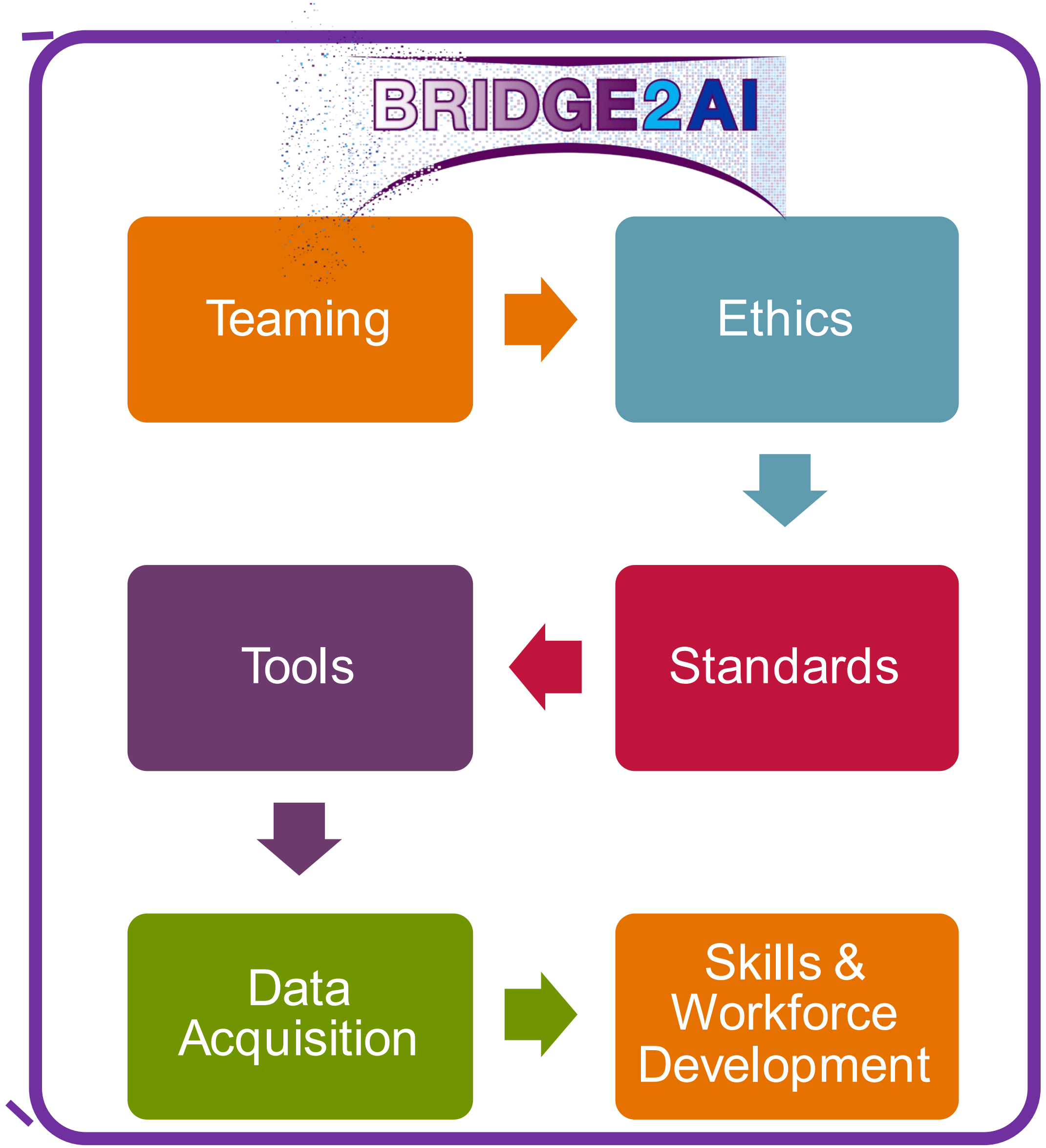


From: Big data hurdles in precision medicine and precision public health, Prosperi et al. BMC Medical Informatics and Decision Making (2018)

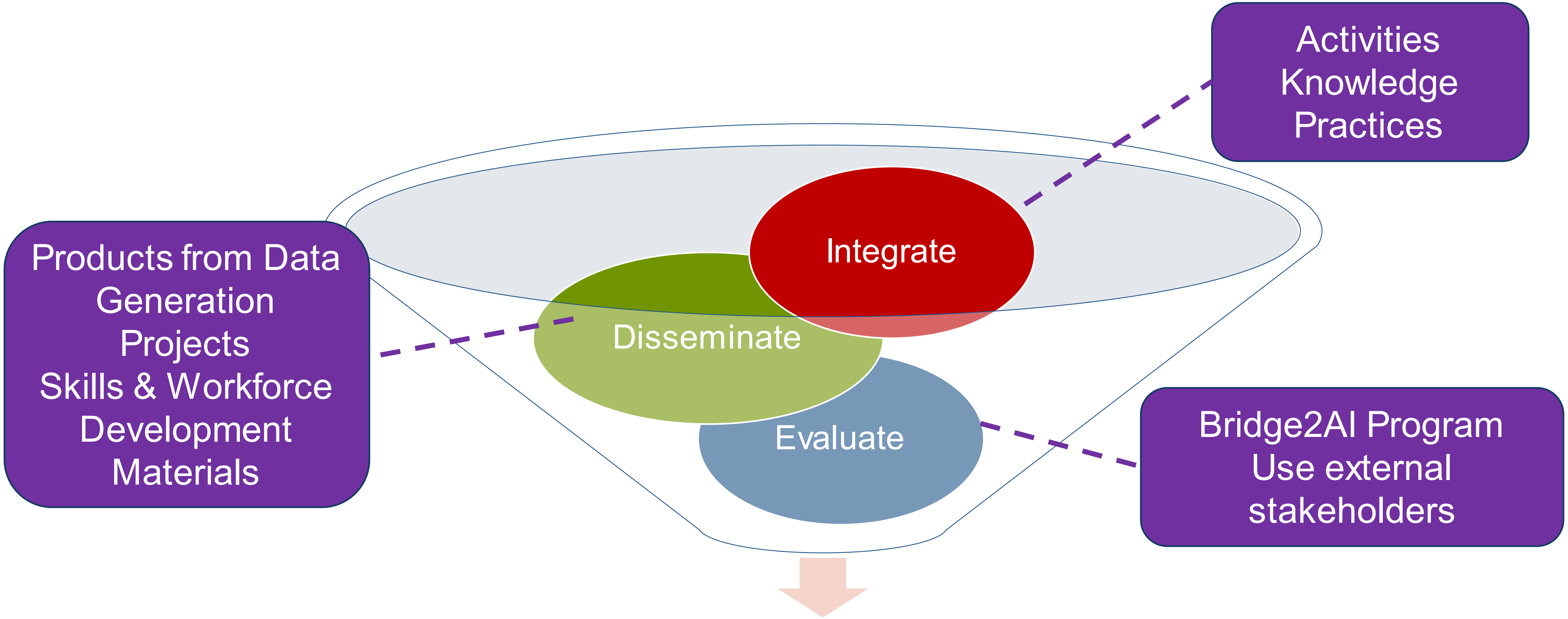
Preparing the Data



- Data Generation Projects (OT2)
  - $96M over 4 years
    5-8 awards
- BRIDGE Center (U54)
  - $8M over 4 years
    1 award



BRIDGE Center



Best Practices for AI/ML in Biomedical and Behavioral Research

FAIR and AI/ML-Ready Data

"Achieving the effective convergence of biomedical data and machine learning requires datasets to be thoughtfully designed from the outset to be valuable for machine learning-based analysis." –NIH ACD Working Group [Report]

What does it mean for data to be AI/ML ready?

- "AI/ML readiness" is not simply formulaic. It requires engagement and feedback from AI/ML applications:
  - Formats are dictated by the AI/ML workflow tools
  - Biomedical applications often require data from multiple sources to be interoperable
  - Other aspects (e.g. representation of information, presence of noise, specificity or uncertainty of labels, and the amount of data) can impact computational and model performance
  - Documentation is also key
- AI/ML-readiness should be guided by a concern for human and clinical impact
  - Requires attention to ethical, legal, and social implications of AI/ML

Collaborations to Make Data FAIR and AI/ML Ready

FY21: NOT-OD-21-094

FY22: NOT-OD-22-067

Support Collaborations to Improve the AI/ML-Readiness of NIH-Supported Data

Artificial intelligence and machine learning (AI/ML) are a collection of data-driven technologies with the potential to significantly advance biomedical research.

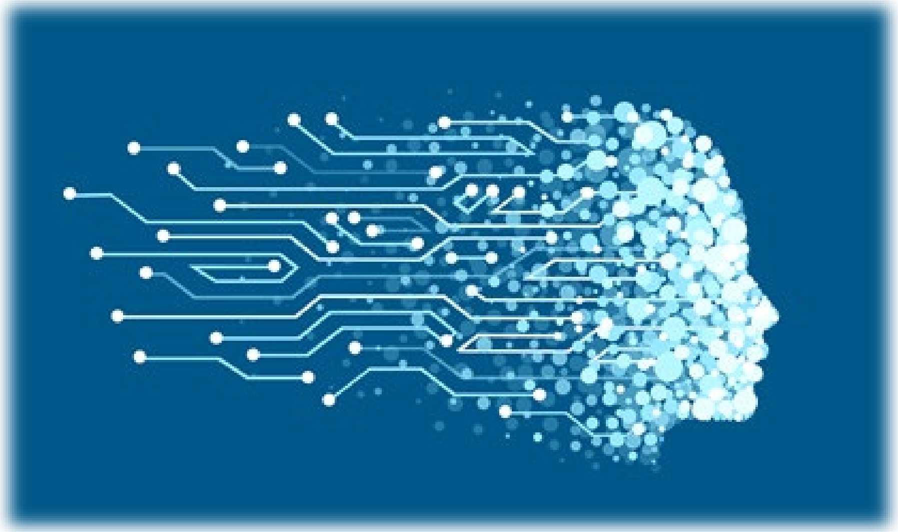
NIH makes a wealth of biomedical data available and reusable to fuel scientific discovery. However, further investment, innovation is needed to ready these data for use for cutting edge AI/ML applications.
To accelerate their development, ODSS supported **collaborations** that brought together expertise in biomedicine, data management, and AI/ML to make NIH-supported data AI-ready for AI/ML analytics.

https://datascience.nih.gov/artificial-intelligence/initiatives/Improving-AI-readiness-of-Existing-Data

Training the Workforce to Make Data FAIR and AI/ML-Ready

FY21: NOT-OD-21-079

ODSS supported the development and implementation of curricular or training activities at the interface of information science, AI/ML, and biomedical sciences to develop the competencies and skills needed to make biomedical data FAIR and AI/ML-ready.



https://datascience.nih.gov/artificial-intelligence/initiatives/Workforce-Gap-Data-Governance-AI

Collaborations to Advance Ethical Use of AI/ML

New in FY22: NOT-OD-22-065

Advancing the Ethical Development and Use of AI/ML in Biomedical and Behavioral Sciences

ODSS will support collaborations that bring together expertise in ethics, biomedicine, data collection, and AI/ML to advance the understanding, tools, metrics, and practices for the ethical development and use of AI/ML in biomedical and behavioral sciences.

These collaborations are intended to generate **new understanding, practices, tools, techniques, metrics, or resources that will aid *others*** in making ethical decisions throughout the development and use of AI/ML, including the collection and generation of data as well as the reuse of data and models by others. Research products developed under this NOSI will be shared and made broadly reusable.

https://datascience.nih.gov/artificial-intelligence/initiatives/ethics-bias-and-transparency-for-people-and-machines

Collaboratively Envisioning AI and Ethics in Biomedical Research

The NIH is interested in bringing together a diverse cross-section of scientists, social scientists, ethicists, advocates, legal scholars, communicators, and artists interested in the social implications of technology to

- **Forge new collaborations among these cross-disciplinary groups**
- **Identify important areas of consideration at the intersection of artificial intelligence (AI) and machine learning (ML), biomedicine, and ethics.**
- **Generate creative strategies to solve ethical dilemmas in biomedical AI/ML**

Collaboratively Envisioning AI and Ethics in Biomedical Research

Micro Lab #1

Dec 15th, 2021, 2-4pm ET

Who are the relevant stakeholders?

Micro Lab #2

Jan 12th, 2022, 2-4pm ET

What are the key opportunities, challenges, and themes?

Micro Lab #3

Jan 26th, 2022, 2-4pm ET

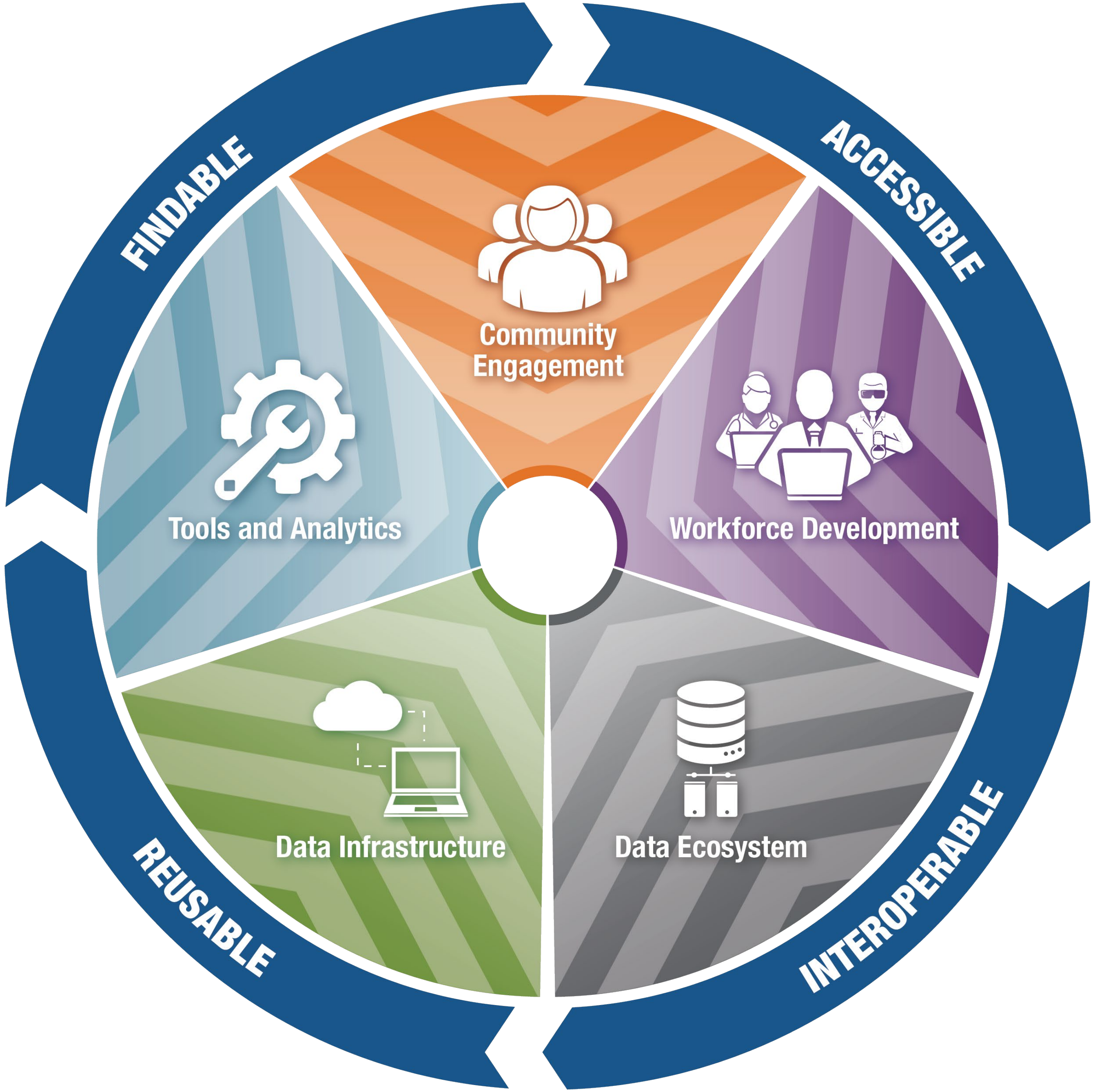
Organizing and understanding opportunity

InnovationLab: A Data Ecosystems Approach to Ethical AI for Biomedical and Behavioral Research

Developing social and technical approaches to defining and implementing ethics across the AI data ecosystem

March 14-18, 2022 from 10:00 AM ET - 5 PM ET.
https://apply.hub.ki/aiandethicsinnovationlab/

Thank you

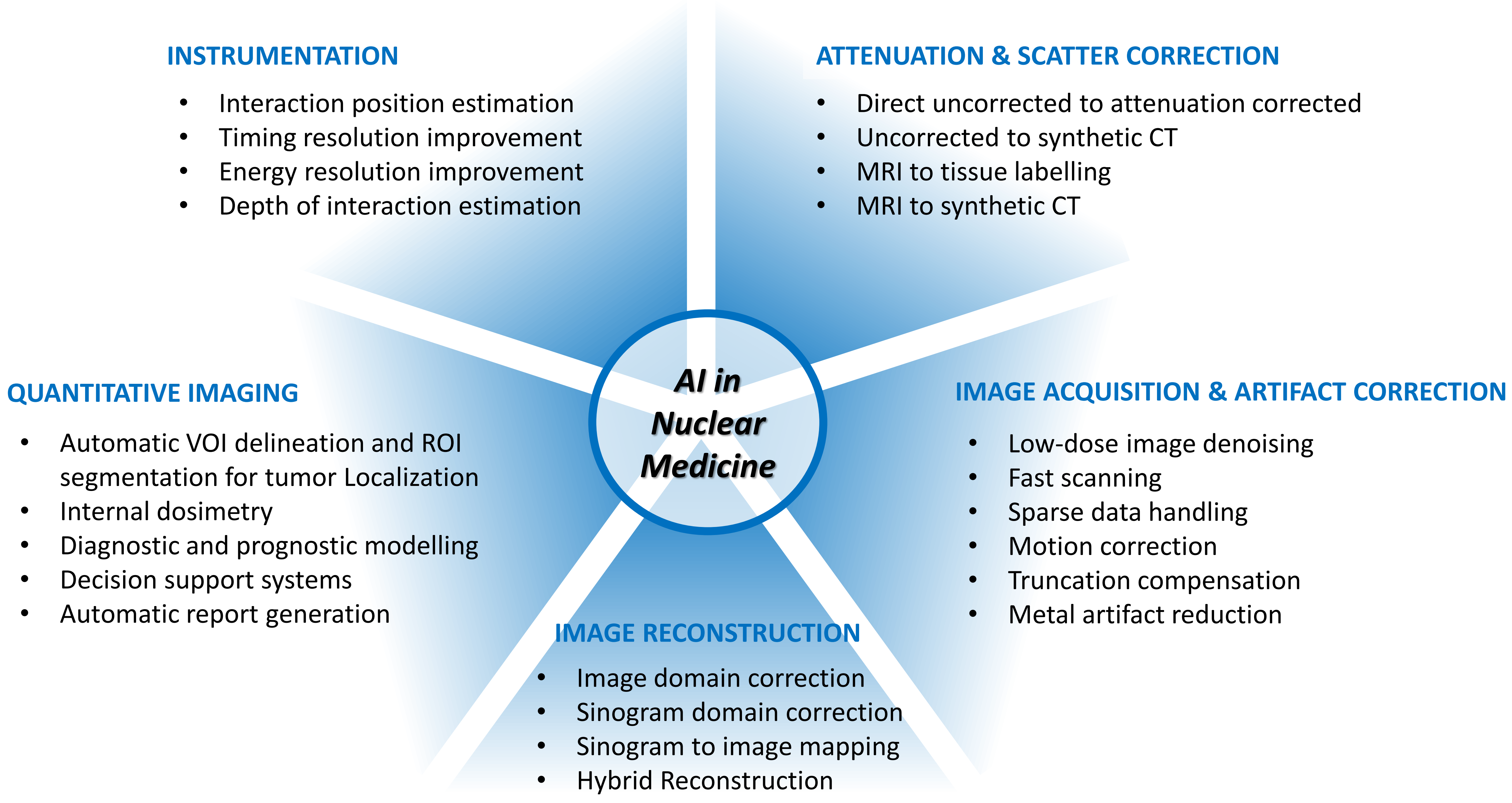


https://datascience.nih.gov/nih-strategic-plan-data-science

AI in Nuclear Medicine
Opportunities, Challenges , and NIBIB Funding

SNMMI AI Summit 2022

Behrouz N. Shabestari, Ph.D.
Director, NIBIB National Technology Centers
Acting Director, Division of Health Informatics Technologies – NIBIB
behrouz.shabestari@nih.gov



Arabi et al, The promise of artificial intelligence and deep learning in PET and SPECT imaging. Physica Medica. 2021

CURRENT WORK ON INSTRUMENTATION

Time of flight estimation

- CNN vs **Leading Edge**
  20% improvement
  (231 ps vs. 185 ps)
- CNN vs **Constant Fraction Discriminator** (CFD)
  23% improvement
  (242 ps vs. 185 ps)

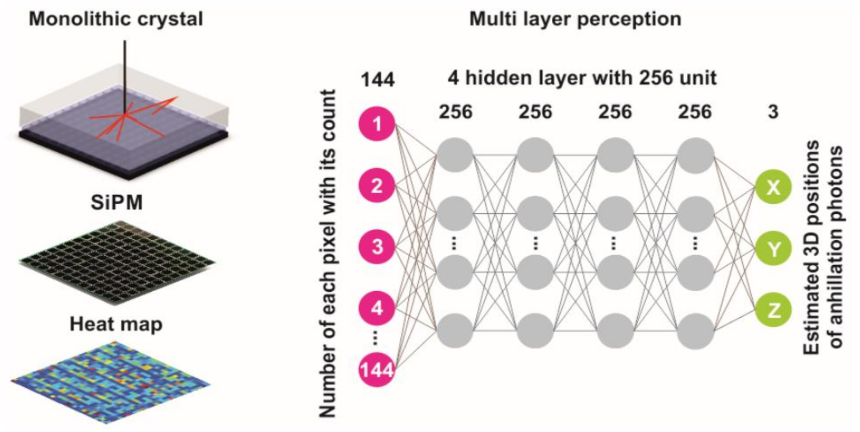
Fixed CNN Configuration

3-Layer	4-Layer	5-Layer	6-Layer	6-Layer (2x feature maps)	7-Layer
conv 2×5 (64) *maxpool 1×2*	conv 2×5 (64)	conv 2×5 (64)	conv 2×5 (64)	conv 2×5 (128)	conv 2×5 (64)
	conv 1×5 (64) *maxpool 1×2*	conv 1×5 (64) *maxpool 1×2*	conv 1×5 (64) *maxpool 1×2*	conv 1×5 (128) *maxpool 1×2*	conv 1×5 (64) *maxpool 1×2*
		conv 1×5 (128) *maxpool 1×2*	conv 1×5 (128)	conv 1×5 (256)	conv 1×5 (128)
			conv 1×5 (128) *maxpool 1×2*	conv 1×5 (256) *maxpool 1×2*	conv 1×5 (128) *maxpool 1×2*
					conv 1×5 (196) *maxpool 1×1*

Berg E, Cherry SR. Using convolutional neural networks to estimate time-of-flight from PET detector waveforms. *Phys Med Biol*. 2018

Position of Interaction

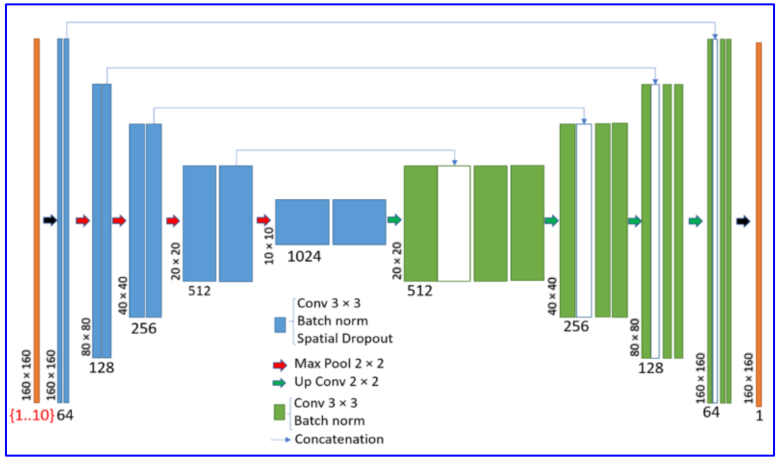
- Multilayer perceptron positioning (MLP): Feedforward Artificial Neural Network (ANN) used for classification problems



Sanaat et al. Depth of Interaction Estimation in a Preclinical PET Scanner Equipped with Monolithic Crystals Coupled to SiPMs Using a Deep Neural Network. Applied Sciences. 2020

Positron Range Correction

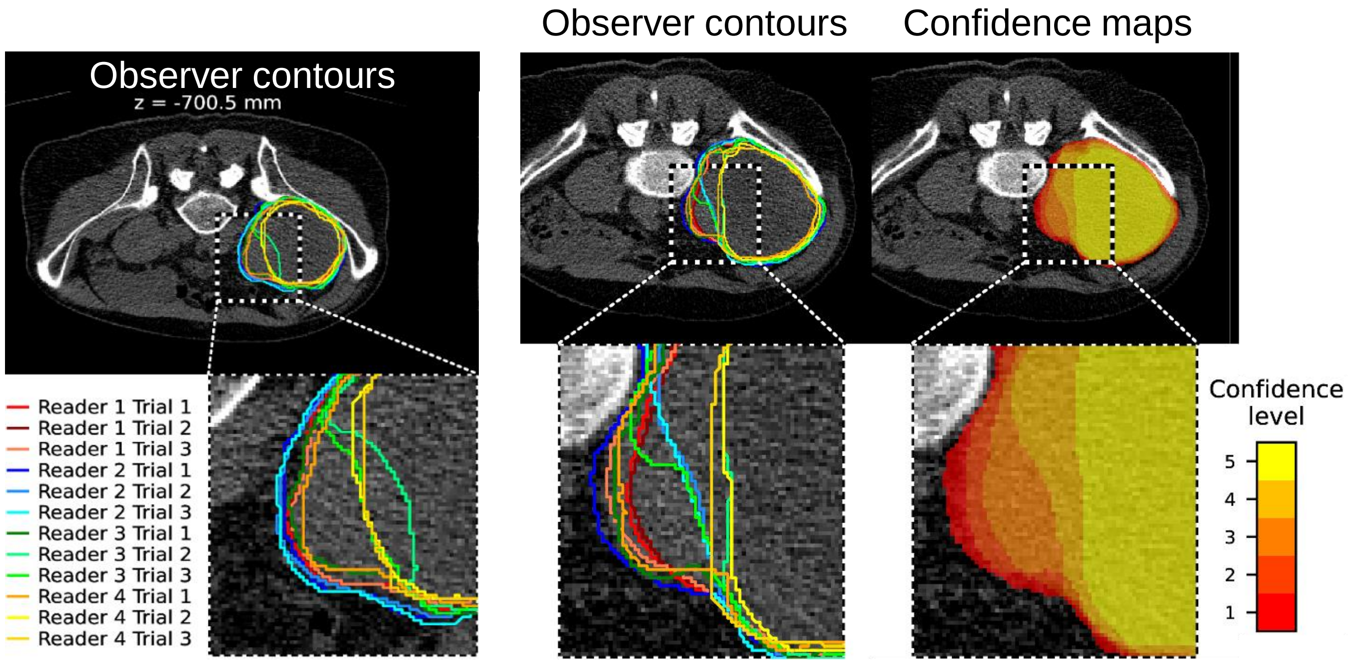
- Modelling the positron range allows for accurate correction
- Spatial resolution is significantly improved



Herraiz et al. Deep-Learning Based Positron Range Correction of PET Images. *Appl. Sci.* 2021

Automatic GTV delineation modeling observer variability

- Gross Tumor Volume (GTV) delineation is a bottleneck in radiation therapy
- Goal: Automatically delineate GTV contours modeling observer variability
- **Challenges**
  - Large amount of data annotations to collect: (4 readers, 3 trials per image, 68 patients)
  - Model variability in deep learning network
- **Opportunities**
  - Predict GTV contour with confidence level
  - Significantly accelerate contouring process
  - Training opportunities for junior radiologists
- **Deep Learning Approach**
  - Learn discrete confidence maps
  - Use modified U-Net structure
- Predicted GTV compared to confidence maps from human observers. Dice score comparable to inter-reader variability.



Comparison	Dice score
Predicted vs. human GTV confidence maps	86.8% (+/- 5.4%)
Inter-observer variability (human)	90.5% (+/- 4.3%)
CTV (from observer vs. predicted GTV)	89.5% (+/- 1.8%)

Gordon Center for Medical Imaging

El Fakhri, P41EB022544 (CMITT), R01CA165221

Utility of other domains

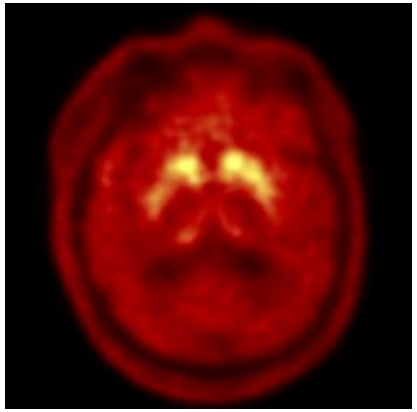
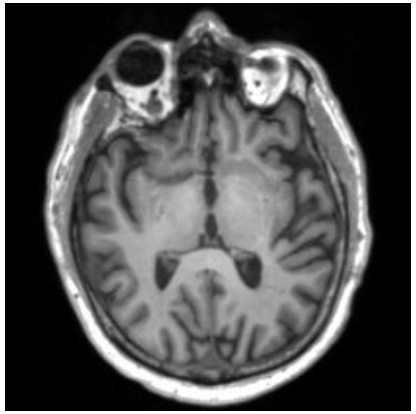
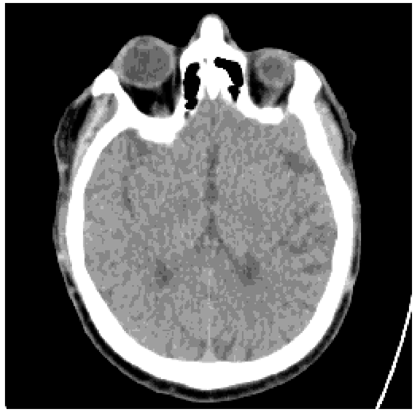
Various domains of medical applications

- Multi-modality: PET / MR / CT

- Multi-tracer PET: FDG, MK6240, FMISO, F-DOPA, …

- Multi-sequence MR: T1, T2, *ASL, MWI, DTI, …*

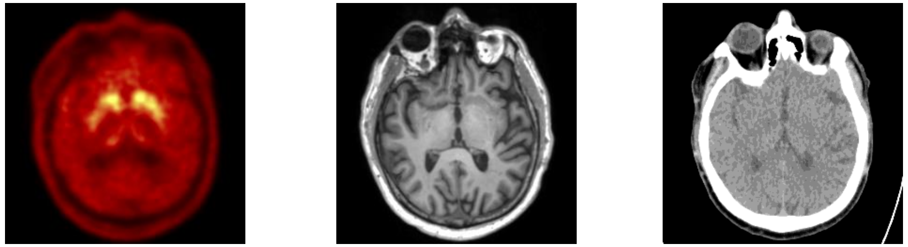
- Images with different scanners / multi-sites

➡ ***How can we utilize other domains?***

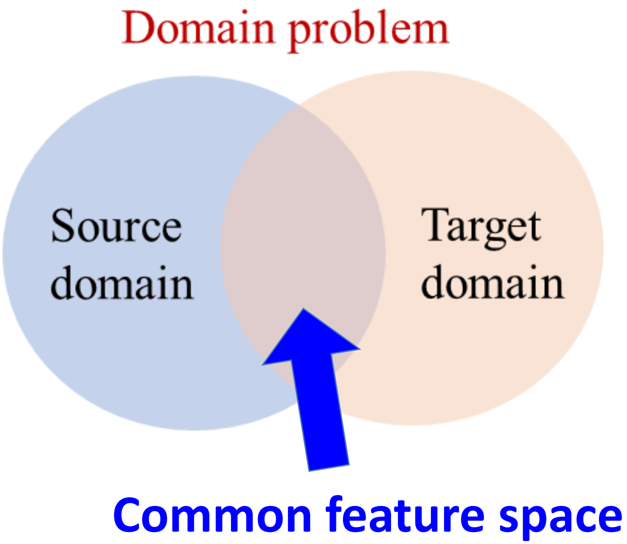
Domain Adaptation/Few Shot Learning

- What is domain? (e.g. brain PET, MR, CT)

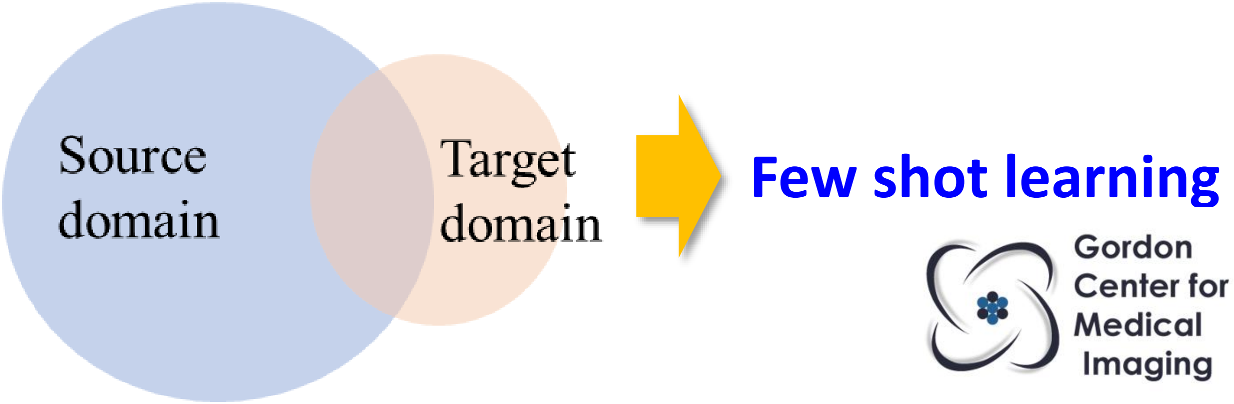


Resolution	★☆☆☆☆	★★★☆☆	★★★★☆
Functional	★★★★☆	★★★☆☆	★☆☆☆☆
Quantification	★★★★☆	★☆☆☆☆	★★★☆☆

Different characteristics for the same brain



- DA has a benefit when target domain has a small dataset



Gordon Center for Medical Imaging

Kim, Li, El Fakhri. P41EB022544

DA for multi-tracer PET image

- Source domain: FDG-PET – sufficient public/internal data
  - ✓ Common features can be utilizable
  - ✓ We assume new tracers may not have sufficient training samples
- Issues of conventional DA
  - ✓ Limitation of data sharing across multi-sites
  - ✓ Inefficient to use large datasets of the source domain
  - ✓ Requires resources & longer training time

- **Goal 1: DA-FSL PET image denoising**
  - ✓ Apply our model to new tracers with insufficient data

Goal 2: Domain adaptation without source data

- ✓ Only <u>trained model</u> in source domain is used

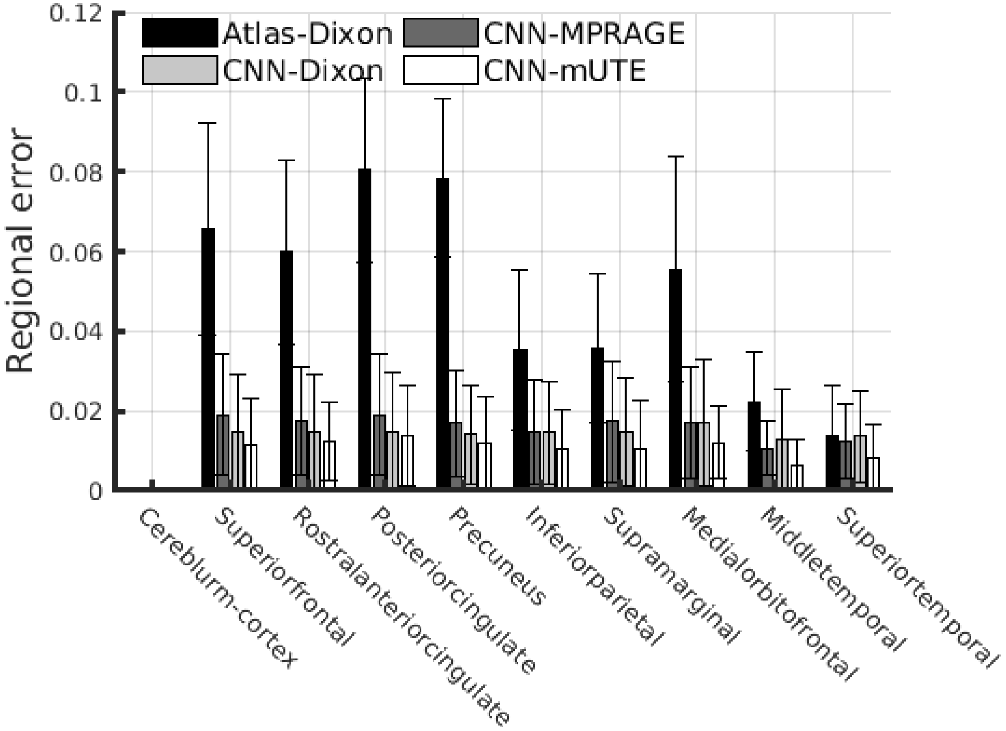
MGH 1811

VE RI TAS

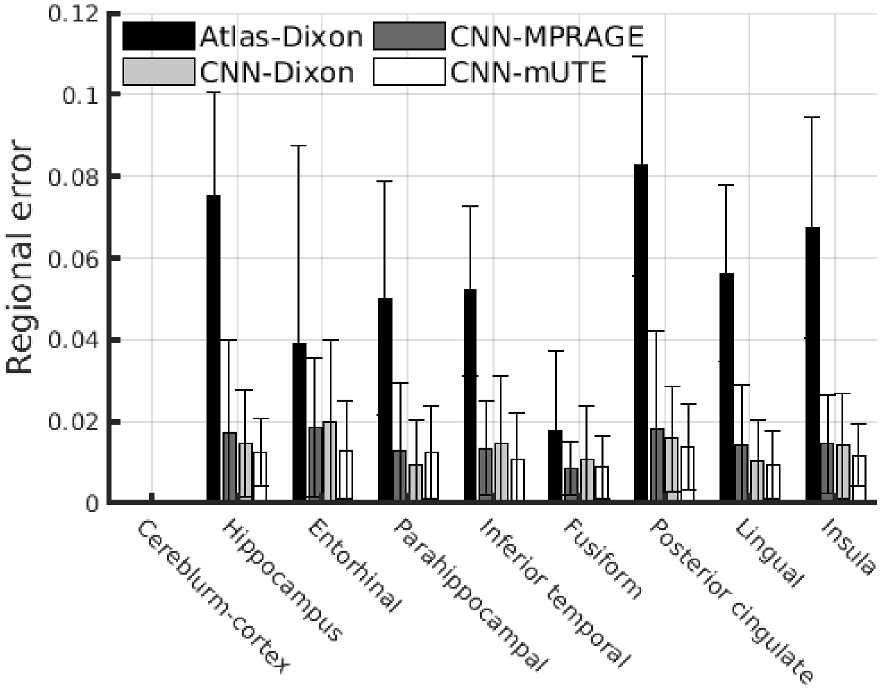
Gordon Center for Medical Imaging

Deep learning-based PET AC for amyloid and tau imaging

SUVR error of amyloid imaging (^{11}C-PiB)



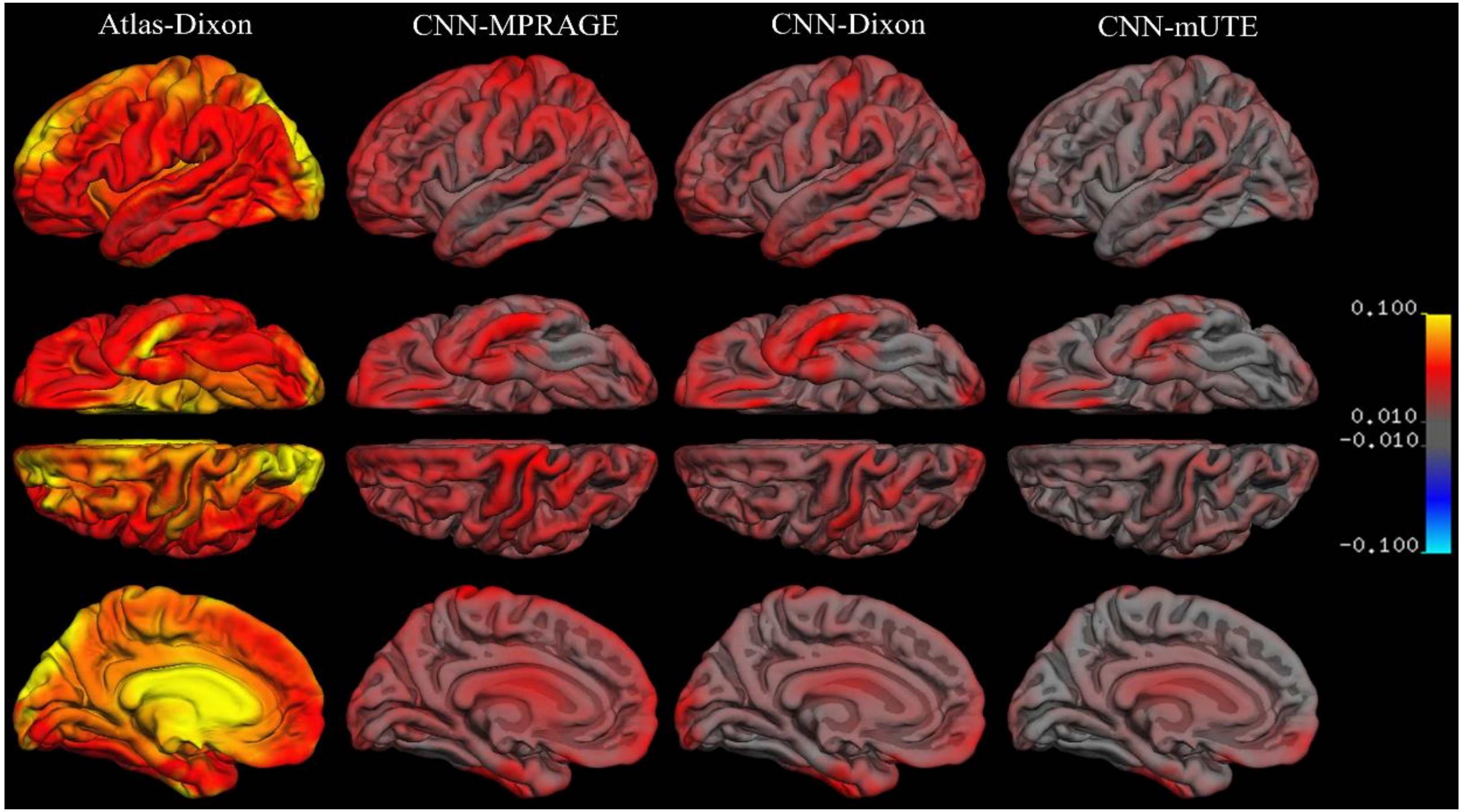
SUVR error of tau imaging (^{18}F-MK-6240)



- The proposed attenuation correction (AC) method by utilizing novel MR-sequence and network-structure designs has the smallest error in amyloid and tau deposition-related regions.

Gordon Center for Medical Imaging

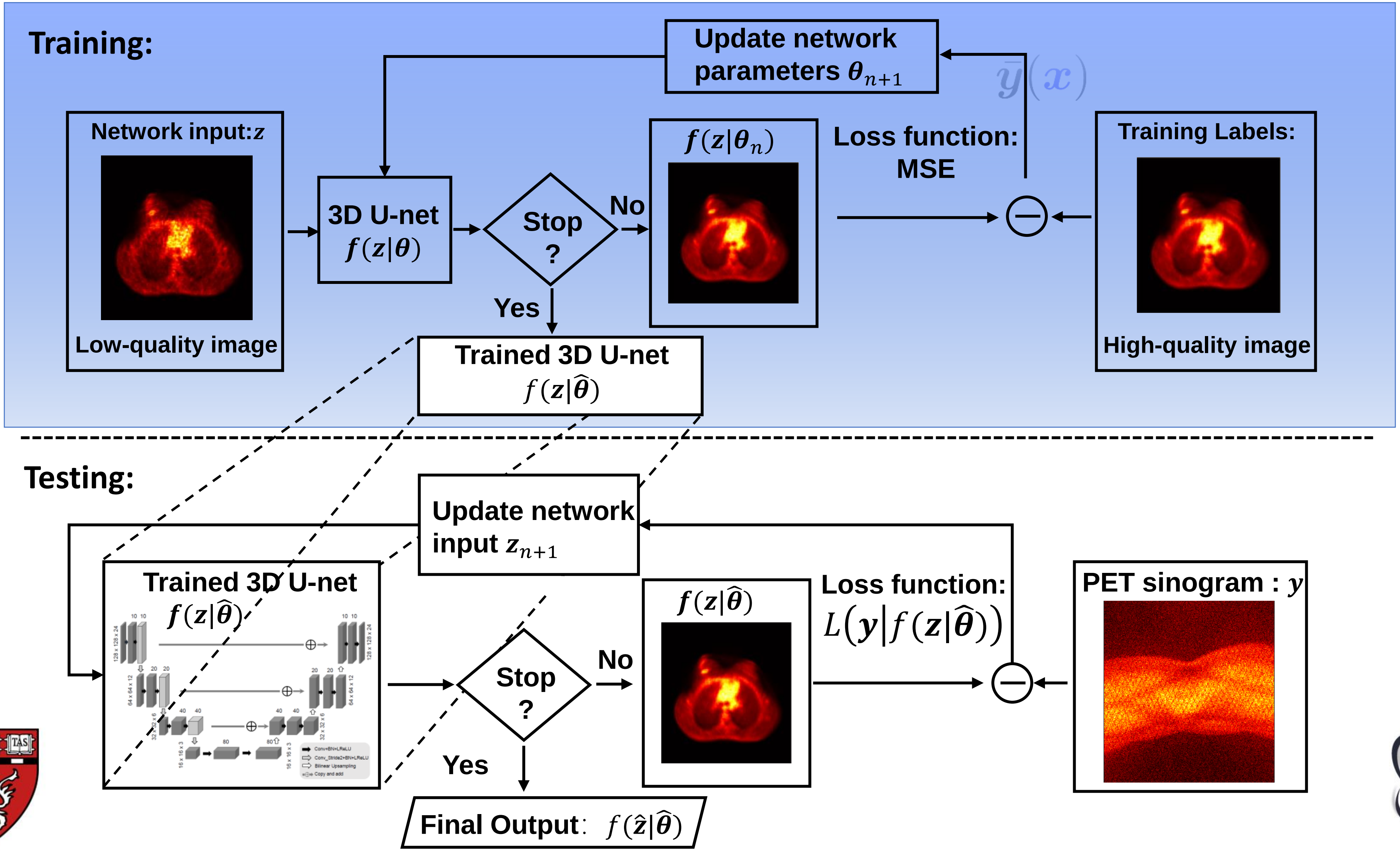
Deep learning-based PET AC for amyloid and tau imaging



- The averaged surface maps of SUVR relative error for different methods. The color map range is from 1% to 10% in magnitude.

Gordon Center for Medical Imaging

Deep learning-based PET image reconstruction

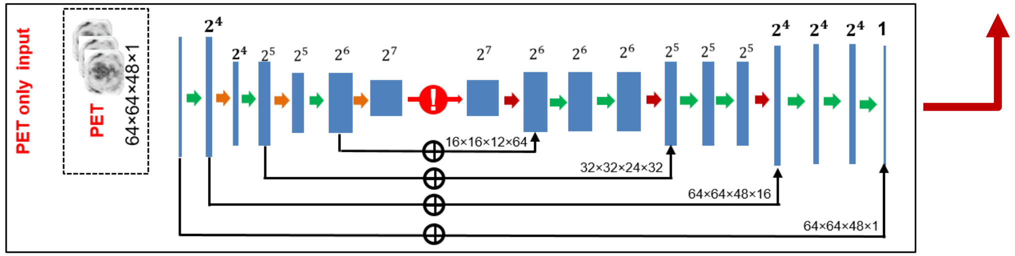


Gordon Center for Medical Imaging

Regularized PET reconstruction using DNN

- Representing the unknown PET image as an output of a pre-trained deep neural network and perform a constrained maximum likelihood estimate:

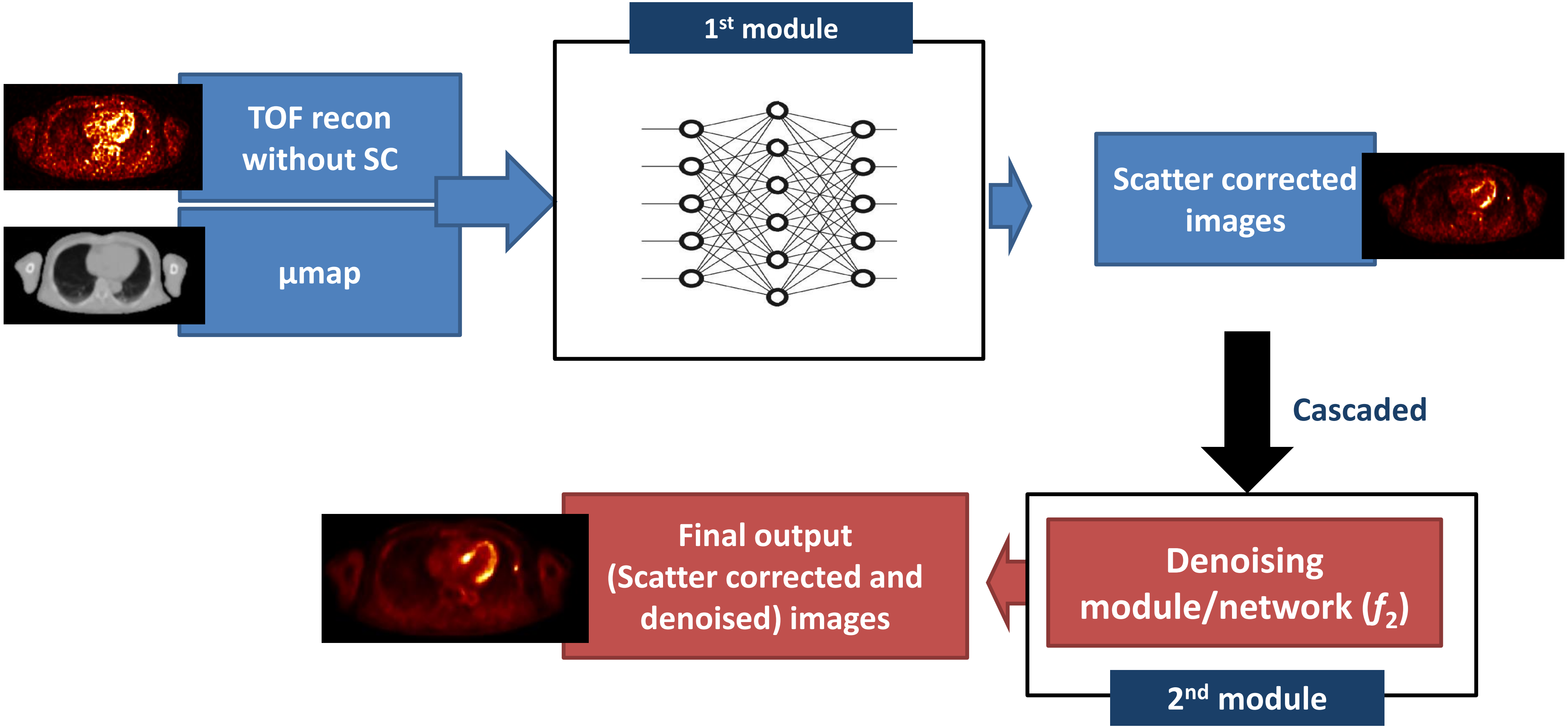
$$\hat{x} = \arg\max_{x} L(y|x)$$



- Both inter-patient information and intra-patient information can be included into the reconstruction by pre-training a DNN using high-quality PET images.

[1] K Gong, J Guan, K Kim, X Zhang, J Yang, Y Seo, G El Fakhri, J Qi, Q Li. Iterative PET image reconstruction using convolutional neural network representation. *IEEE TMI*, 2018
[2] Z Xie, X Zhang, T Li, W Qi, E Asma, J Qi. Generative adversarial network based regularized image reconstruction for PET. Phys Med Biol. 2020
[3] Z Xie et al, **Anatomically aided PET image reconstruction with deep neural network**. Medical Physics, 2021.

Cascaded a 3D CNN denoising module



Main Challenges

- Lack of a very large amount and volume of high-quality (clinical) training data-images.
  - clinical 3D data are typically very large (as compared to reconstructed images), and are not always stored on the clinical systems

  *** Potential solution: **using simulated data based on physical imaging models to pre-train network and fine tune using real data**
- Need harmonized data and images

  *** Potential solutions: **artificial and virtual data for training – use of phantom data!**
- Need of a large number of data-image pairs for proper training of deep networks (with huge amount of connections/parameters)

  *** Potential solutions: **data augmentation and transfer learning techniques**
- Training and generalizing the networks across sites - difficulties to exchange data
  - proprietary data formats by different manufacturers

    *** Potential solutions**: federated approaches – training of identical networks at different centers and sharing only the trained network parameters**

Opportunities

- Ability to work with a very low count data to provide diagnostic quality images
- Ability to work with imperfect and contaminated data
- Ultra fast, near real-time, reconstructions directly from data (especially important for motion and dynamic studies with many time frames, interventional procedures, etc.)
- In quantitative imaging, deep learning-based methods provide faster alternatives with high accuracy and can also perform attenuation correction simultaneously
- Deep neural networks provide new ways to design the regularization function
- Promise in **development of novel PET tracers** and cardiac-specific postprocessing techniques using artificial intelligence
- Significant opportunities to reduce noise and improve reconstruction

Funding Opportunities at NIBIB

Bioengineering Partnership with Industry (BPI) (U01)

PAR-22-123

NIBIB Notice of Intent to Publish a Funding Opportunity Announcement

- **Participating ICs: NIBIB,** NIA**, NEI,** NCI
- Posted Date: March 08, 2022
- **Application Due Date: May 26, 2022**

- **Purpose:** The use of engineering principles to drive development, speed the adaptation, and establish tools and technologies as robust, well-characterized solutions that fulfill an unmet need and to encourage applications to:

    1. establish a robust engineering solution to a problem in biomedical research or the practice of medicine;
    2. develop a strategic alliance of multi-disciplinary partners based on a well-defined leadership plan; and
    3. realize a specific endpoint within 5-10 years with a detailed plan, timeline and quantitative milestones.

- A Key Requirement: BPI applications must include at least **1** academic and **1** industrial organization.
- The areas of research: must be consistent with the missions of the IC's participating in the BPI

Bioengineering Partnership with Industry (BPI) (U01)

PAR-22-123

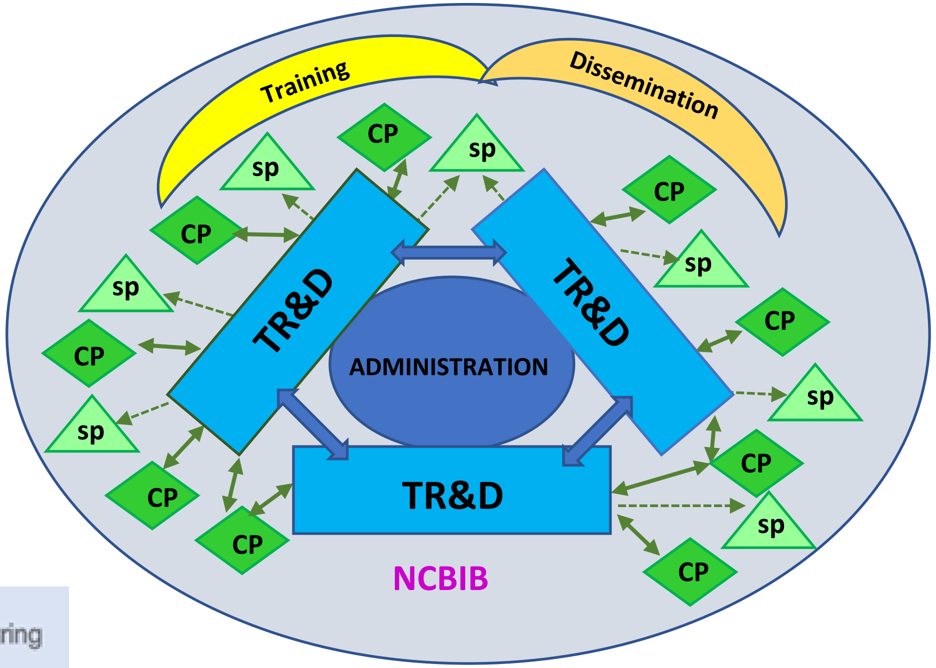
- U01 (cooperative agreement) mechanism
- Clinical applications optional, but encouraged
- Milestones and deliverables / interim reports
- Duration of 5 years
- One competitive renewal
- Budget – applications requesting ≥$500k/year require IC approval
- Require an Industrial Partnership

National Centers for Biomedical Imaging and Bioengineering (NCBIB) - P41

- **Strong foundation of Technology Research & Development (TR&D) Projects**
  - technology development, not mechanistic research; within NIBIB mission
  - **national/international impact -- uniqueness**
  - innovative, cutting-edge, responsive to current challenges in the field
  - complex, multidisciplinary – synergy among TR&Ds
  - high-risk test beds leading to practical tools
- **Driven by needs of the field through robust Collaborative Projects (CP)**
  - **dynamic, iterative push-pull relationships**
- **Deploying results via Service Projects (SP)**
  - **geographically diverse**
  - technology push (using tools not available elsewhere) -- exploit more mature capabilities of the Center

Seamless Oversight

- Senior scientist as PI
- experienced TR&D Leaders
- External Advisory Board
- Institutional Support



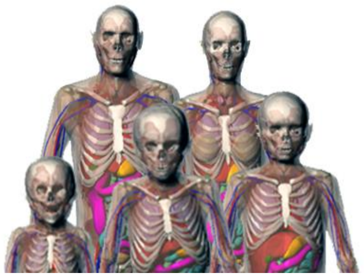
Training and Dissemination

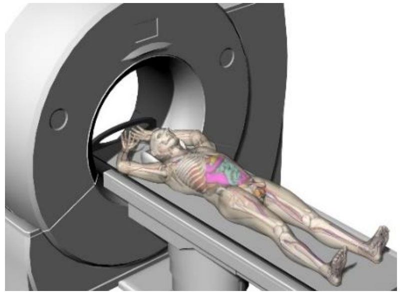
- Committed to training practitioners
- Aggressive dissemination
  - research papers, reviews
  - patents
  - presentations, workshops
  - website(s), newsletters
  - public outreach

P41 Centers: Center for Virtual Imaging Trials

virtual trial

virtual population | virtual acquisition | virtual interpretation





Duke Radiology
Duke University School of Medicine

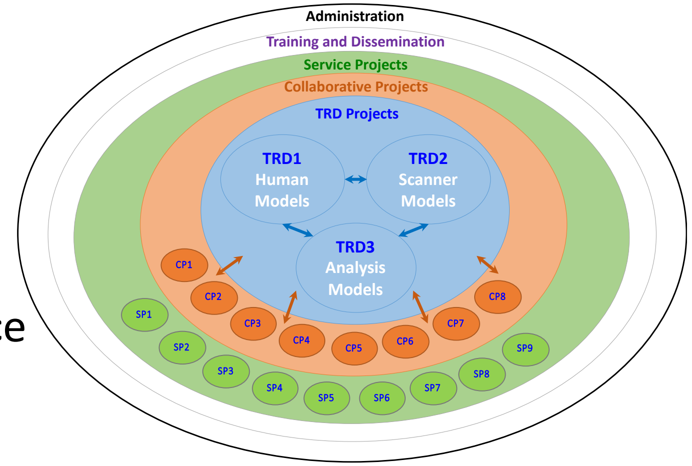
a new experimental paradigm in medicine

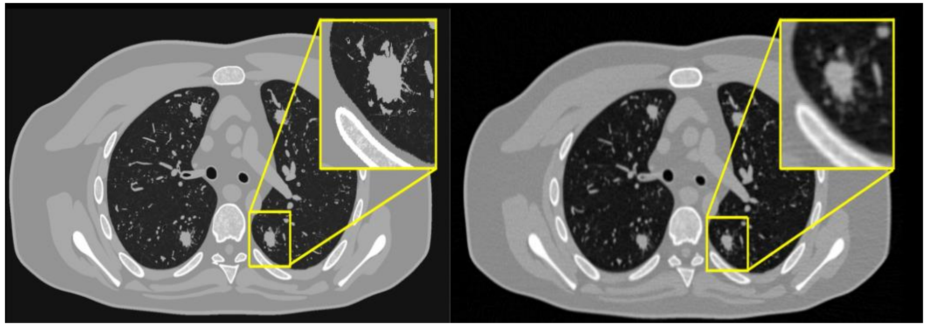
https://deckard.duhs.duke.edu/cvit/

A national center to develop and provide a virtual platform to assess the clinical performance of medical imaging systems from design to use

Serving a broad coalition of
- Academia: Stanford, Yale, Harvard, …
- Industry: GE, Siemens, HeartFlow,
- Government: NIH, NASA, …

A platform for new science
A new method to test and optimize practice



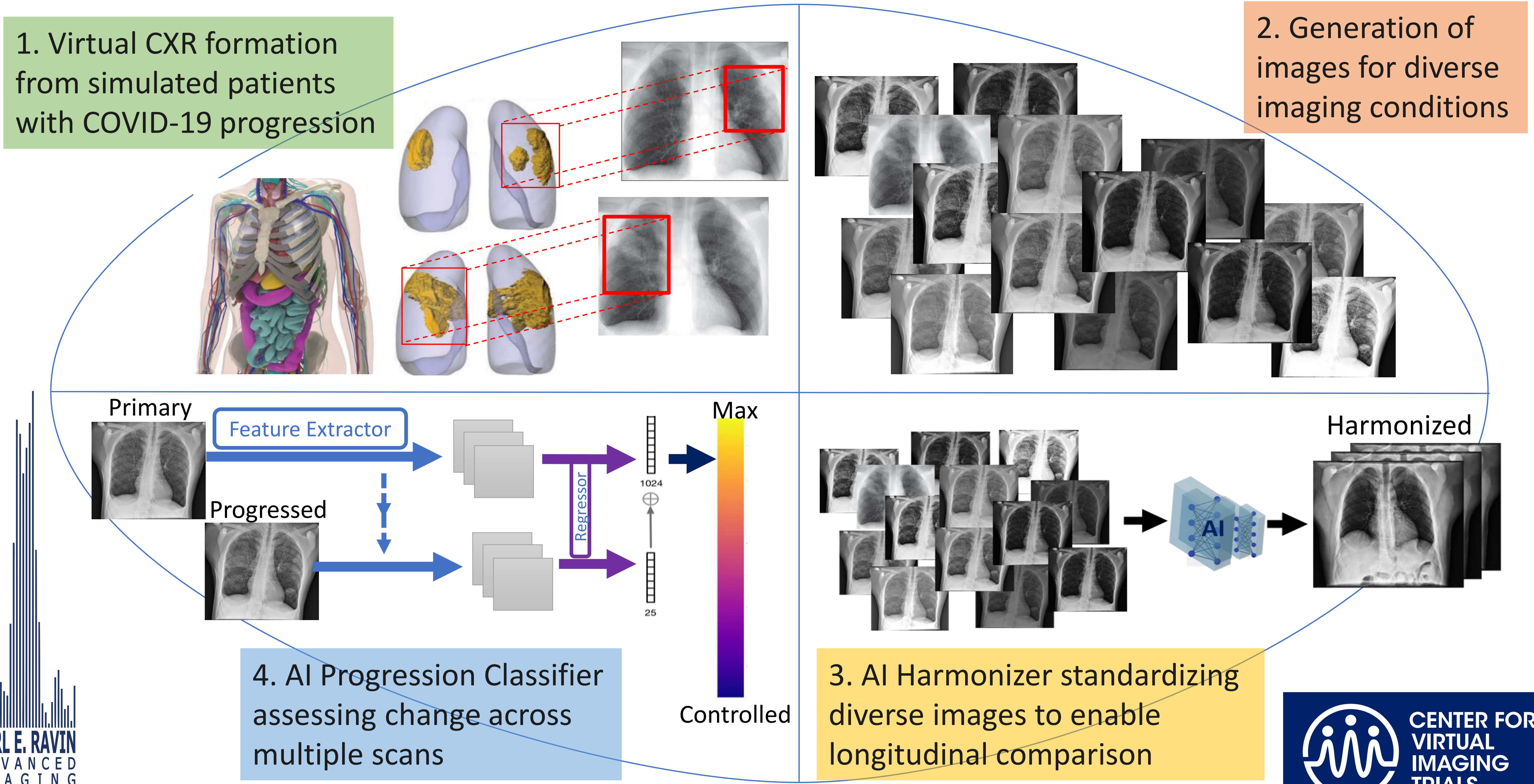


Ground truth **Simulated CT image**

Siemens Flash, 120 kV, pitch of 1, "body" filter

SAMEI, EHSAN, Duke University, 5 P41 EB028744-02

Virtual Clinical Trial Can Enable Assessment and Management of COVID PASC



CARL E. RAVIN ADVANCED IMAGING LABORATORIES

CENTER FOR VIRTUAL IMAGING TRIALS

Stephen I. Katz Early-Stage Investigator Research Project Grant

- **R01 (Clinical Trial Not Allowed)**
- **New R01 FOA (PAR-21-038) Release Date: November 9, 2020**
  - Standard Submission Dates

  Google: "NIH Katz award"
- **Specifically for Early-Stage Investigators**
  - Up to 5-years may be requested
  - Must **not** include preliminary data
- **Encourages:**
  - An innovative project that represents **a change in research direction**
  - Applications must include a separate attachment describing the change in research direction.
  - Early-stage developmental ideas that promise transformation
  - High-risk/High-reward projects

Stephen I Katz, M.D., Ph.D.
Director NIAMS 1995-2018

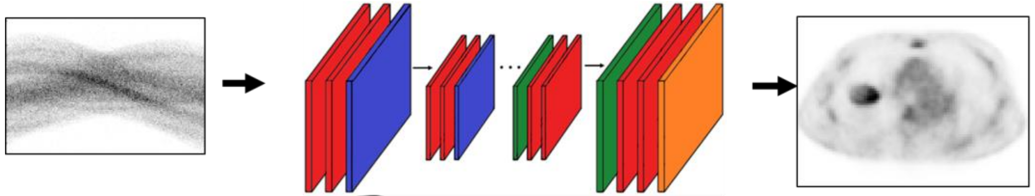
Acknowledgement

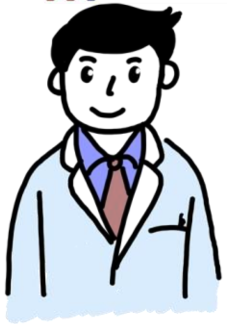
Georges El Fakhri, Ph.D. -- MGH
Jinyi Qi – UC Davis
Samuel Matej, Ph.D. – Upenn
Tatjana Atanasijevic, Ph.D. -- NIBIB

Thank You

Selected Funding opportunities related to advanced imaging AI at NCI

- "Integration of Imaging and Fluid-Based Tumor Monitoring in Cancer Therapy" **PAR-21-290 (R01)**
- "Molecular Imaging of Inflammation in Cancer" **PAR-21-294 (R01)**
- **Notice of Special Interest (NOSI):** Translation of Quantitative Imaging tools and Methods for the Academic Industrial Partnership (AIP) **NOT-CA-21-032**
- **Notice of Special Interest (NOSI):** Advancing the development of tumor site-activated small molecules **NOT-CA-21-101**
- **Notice of Special Interest (NOSI):** Research on Interprofessional Teamwork and Coordination During Cancer Diagnosis and Treatment; **NOT-CA-22-014**
- **Notice of Special Interest (NOSI):** Validation of Digital Health and **Artificial Intelligence** Tools for Improved Assessment in Epidemiological, Clinical, and Intervention Research **NOT-CA-22-037**





Is this AI output reliable?

Evaluating AI algorithms for nuclear medicine: Ongoing efforts and the road ahead

Abhinav K. Jha, PhD
Department of Biomedical Engineering
Mallinckrodt Institute of Radiology
<u>**SNMMI AI Taskforce Evaluation Team lead**</u>

SNMMI Value Initiative
SOCIETY OF NUCLEAR MEDICINE & MOLECULAR IMAGING

SNMMI AI Summit 2022

Washington University in St. Louis
McKelvey School of Engineering
School of Medicine, Mallinckrodt Institute of Radiology

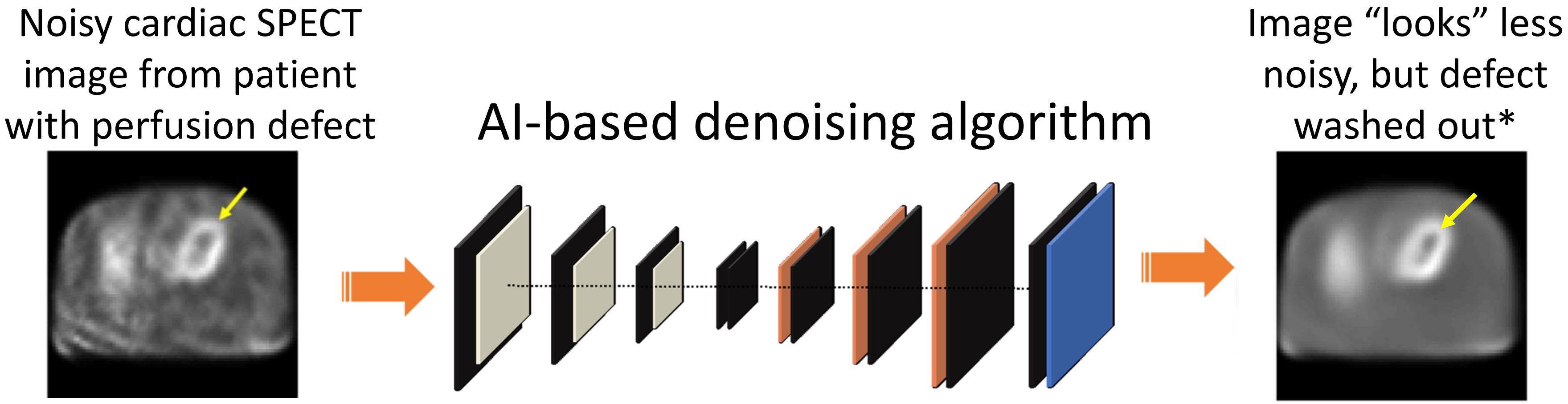
Outline

- Challenges in evaluation of AI algorithms
- Efforts of the SNMMI AI taskforce evaluation team
- Other ongoing efforts towards evaluating AI algorithms for nuclear medicine
- Road ahead: Some important needs
- A wishlist ☺

Introduction

- AI algorithms are showing significant promise in multiple aspects of nuclear medicine
- For clinical translation of AI algorithms, rigorous evaluation is imperative
- AI algorithms learn rules from analysis of training data. Thus:
  - Their performance depends heavily on the training data
  - Output often not interpretable and can be unpredictable
- This leads to several challenges that the evaluation strategy should be able to address

Challenge: Task-agnostic evaluation may not reflect performance on clinical tasks



Evaluation using task-agnostic metrics (root mean square error for reconstruction/denoising and Dice scores for segmentation) may not correlate with performance on clinical tasks*

Evaluation should assess performance on clinical tasks

*Yu et al, J. Nuc. Med. 2019
*Yang et al, Rad. AI, 2020

Challenge: Generalizability

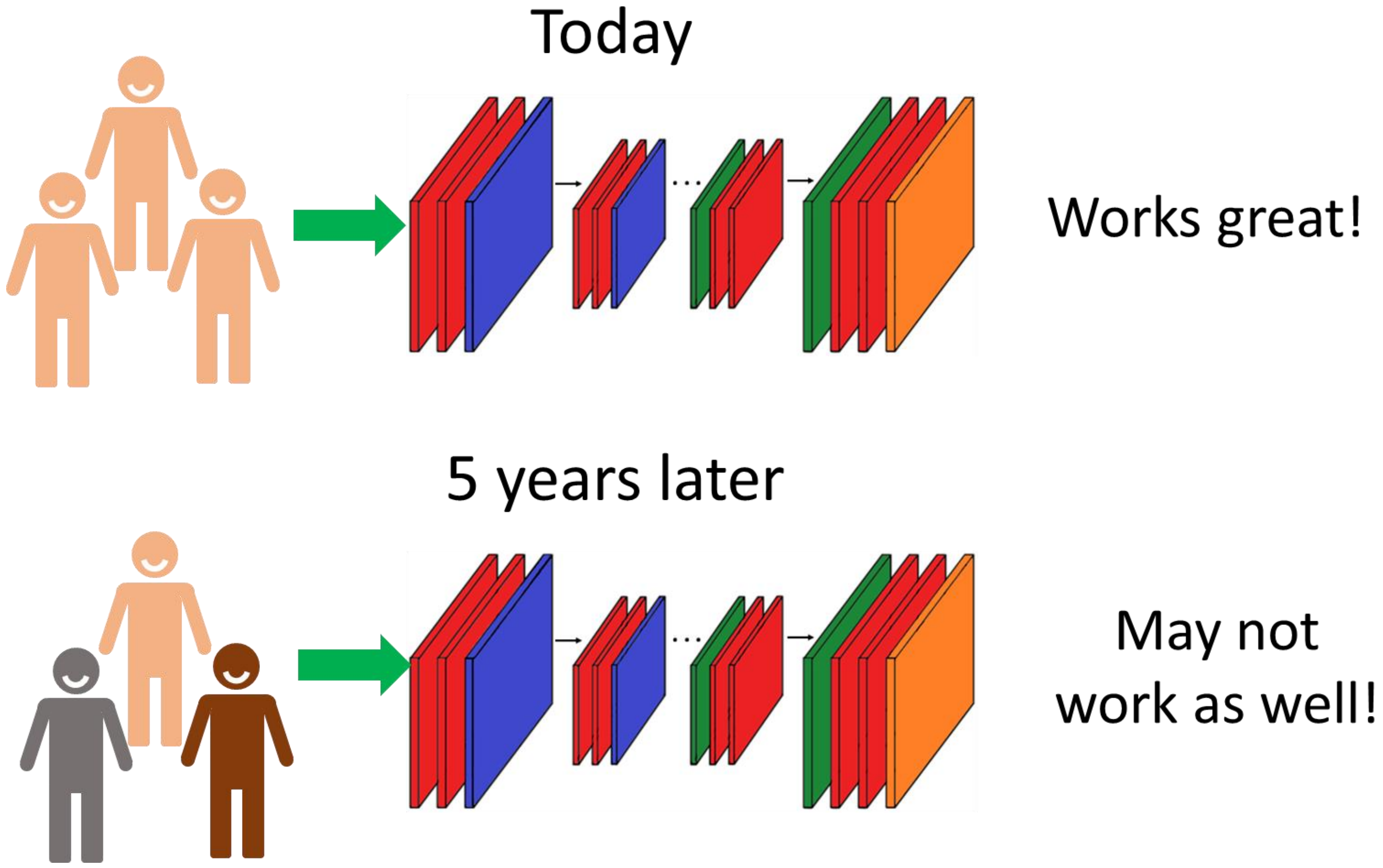
Trained at this center and
worked great!

Oops!

Evaluation should characterize the generalizability of AI methods

Zech et al, PLOS Med, 2018
Gianfranceso et al, JAMA Intern Med. 2018
Noor et al, BMJ 2020

Challenge: Data drift



Evaluation should assess if the method is performing reliably in a post-deployment setting

A major ongoing effort: SNMMI AI Taskforce Evaluation Team

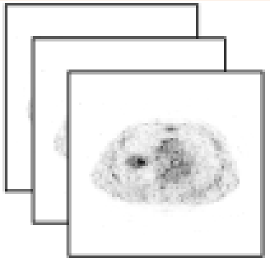
Team consisting of nuclear medicine physicists, computational imaging scientists, physicians, statisticians, representatives from the industry and from regulatory agencies

A key recommendation from the taskforce: The claim

An evaluation study for an AI algorithm should produce an accompanying claim consisting of the following components

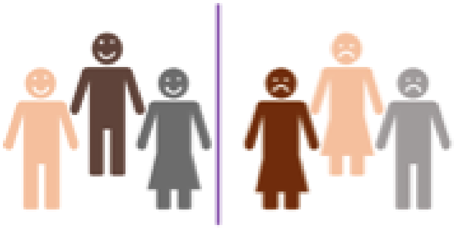
1. Definition of clinical task

- Classify
- Quantify
- Jointly classify & quantify



2. Patient population

- Should be representative of target population
- Demographics including sex, age, ethnicity should be stated



3. Imaging process

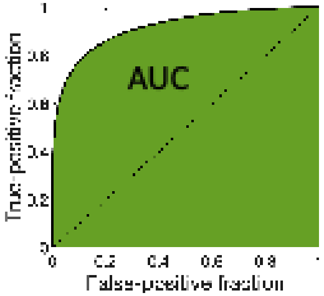
- Imaging system(s)
- image-acquisition protocol(s)
- Single/multi center

4. Strategy to extract task-specific information

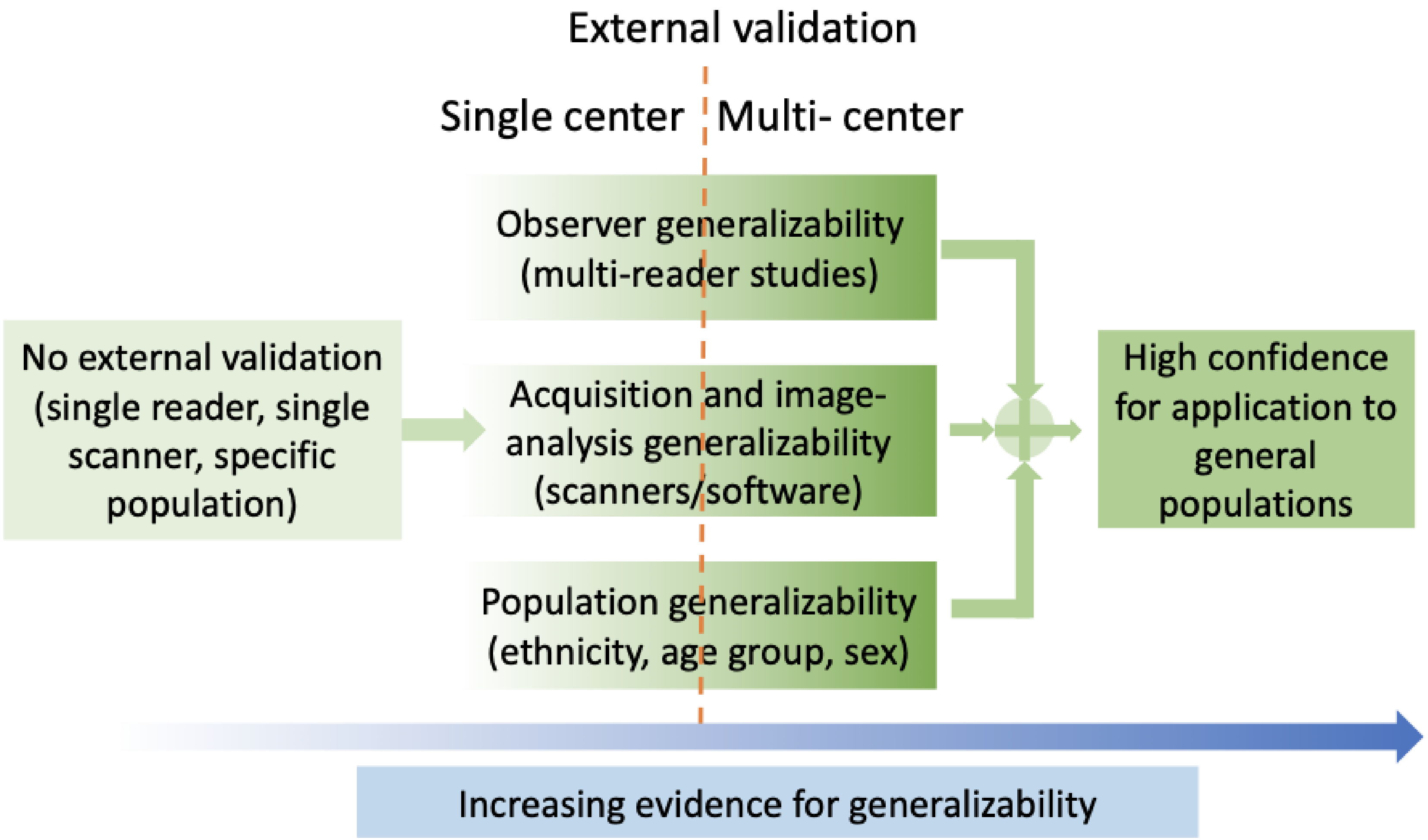
- Detection: Human/model observers
- Quantification: Bayesian and frequentist
- Single/multi reader studies

5. Figure of merit to quantify performance

- Detection: ROC analysis
- Quantification: Ensemble bias and variance, EMSE, bias/variance profiles
- Jointly detect and quantify: LROC, EROC, FROC, LROC



The claim will inherently quantify the generalizability of the AI algorithm



The task force proposes an evaluation framework

Clinically effective post deployment

Method works robustly with populations.

Efficacy in making clinical decisions

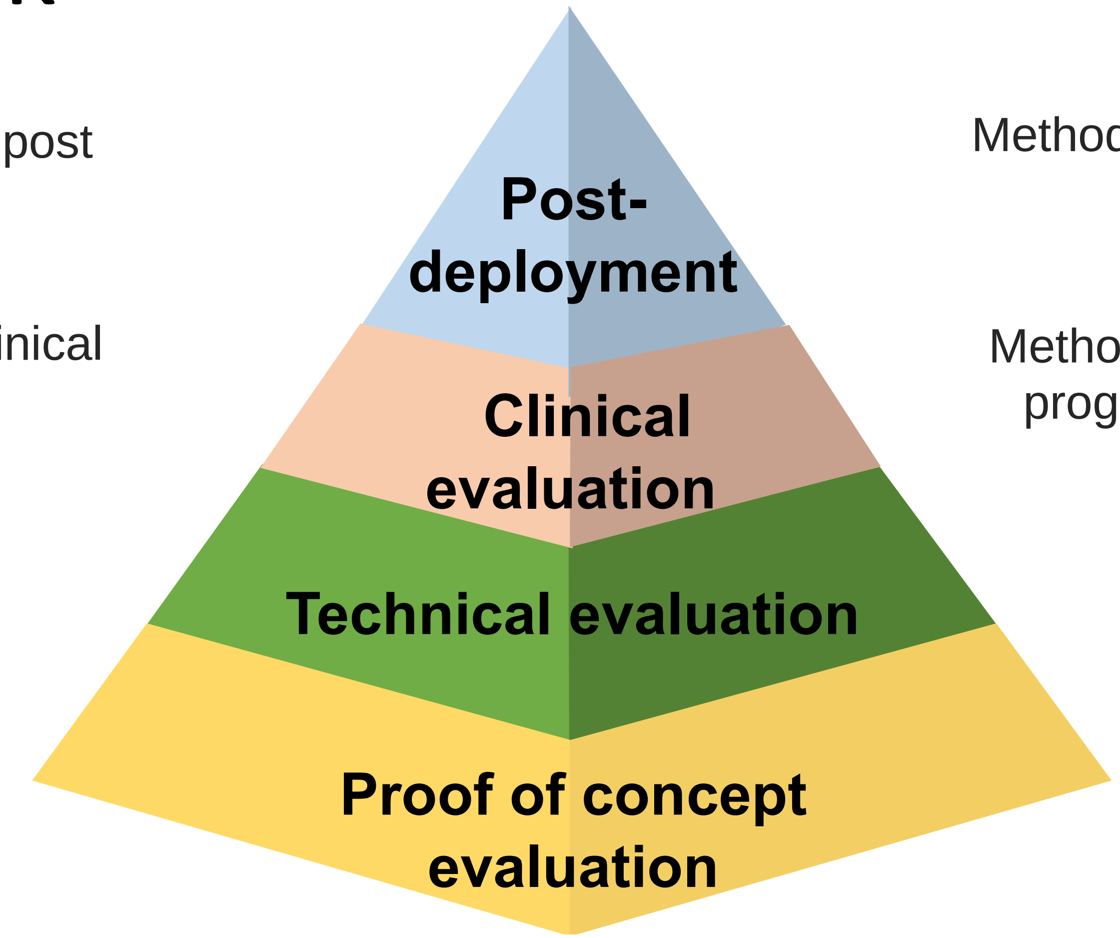
Method-derived MTV values can prognosticate patient response

Efficacy on task-specific technical aspects

Method yields accurate and precise MTV values

Method shows promise.

AI-based segmentation method evaluated with Dice scores

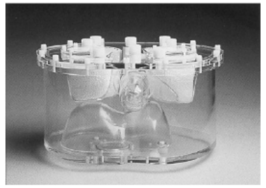


This framework will guide AI developers conduct the evaluation study that provides evidence to support their intended claim

Elements of study design for each class of evaluation

1. Determine study type

- Simulations
- Phantoms
- Clinical studies
  - Retrospective/prospective
  - Observational/interventional

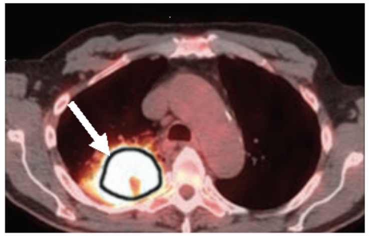


2. Data collection

- Representative
- No overlap between test and training data
- External validation for clinical/technical evaluation

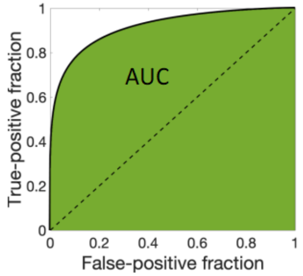
3. Define reference standard

- Quality may vary based on the claim
- Single/multi-reader studies
- Clinical outcomes



4. Choosing figure of merit

- Proof-of-concept eval.: Show clinical promise
- Technical eval.: Accuracy, Precision
- Clinical eval. Show clinical utility



The taskforce is providing the RELAINCE (**R**ecommendations for **E**valuation of **AI** in **N**u**c**lear Medicin**e**) guidelines for each element of study design

The RELAINCE guidelines

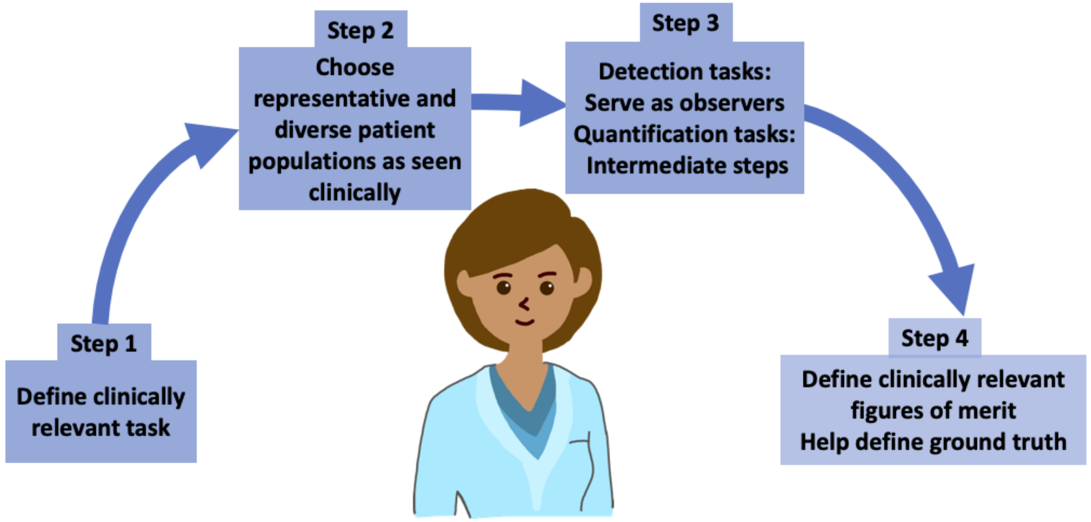
- Provide best practices for evaluation in each element of study design
- Proposed for each class of evaluation
- More details in forthcoming paper*

*Jha et al, AI in Nuclear Medicine: Best practices for Evaluation, J. Nuc. Med., accepted with minor revisions

A recent effort towards implementing some of the RELAINCE guidelines*

Objective Task-Based Evaluation of Artificial Intelligence-Based Medical Imaging Methods:
Framework, Strategies, and Role of the Physician

Abhinav K. Jha, PhD[a,*], Kyle J. Myers, PhD[b], Nancy A. Obuchowski, PhD[c], Ziping Liu, BS[d], Md Ashequr Rahman, BS[d], Babak Saboury, MD, MPH, DABR, DABNM[e], Arman Rahmim, PhD, DABSNM[f], Barry A. Siegel, MD[g]



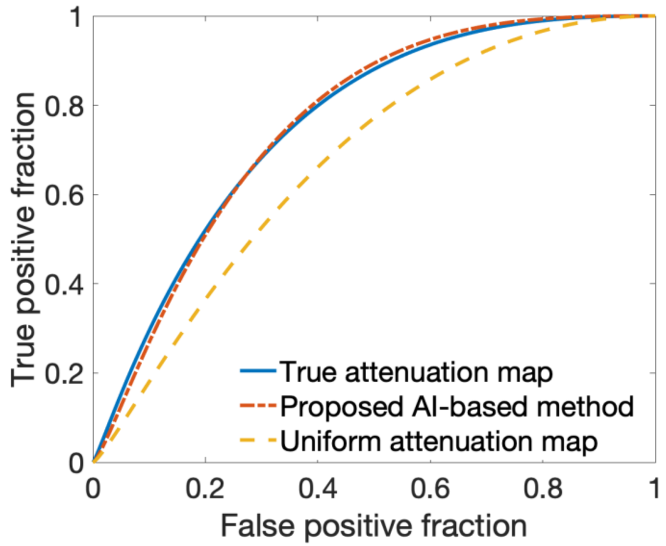
Paper provides the tools to implement some of RELAINCE guidelines in context of PET

*Jha et al, PET Clinics, 2021

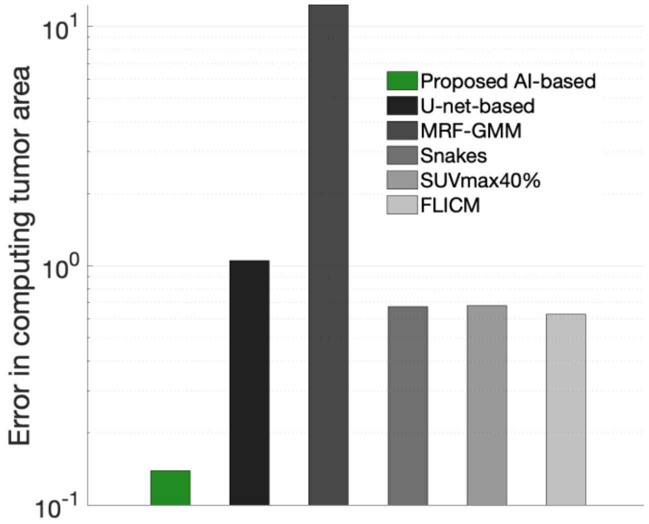
Other ongoing efforts

- Evaluation team, AI-dosimetry taskforce: Goal is to develop guidelines for task-based evaluation of AI methods for image-based dosimetry
- Nuclear-medicine data standardization initiative

Evaluation of AI-based transmission-less attenuation compensation method for cardiac SPECT on defect-detection task using a virtual clinical trial
(Yu et al, SPIE Proc. 2021)

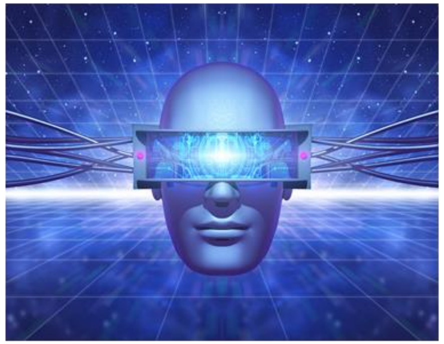


Evaluation of an AI-based PET segmentation method for oncological PET on quantification task using ACRIN 6668 multi-center clinical trial data
(Liu et al, Phys. Med. Biol. 2021, highlighted on NIBIB website)



The road ahead: Vision

AI algorithms



+

Clinical task-based evaluation

=

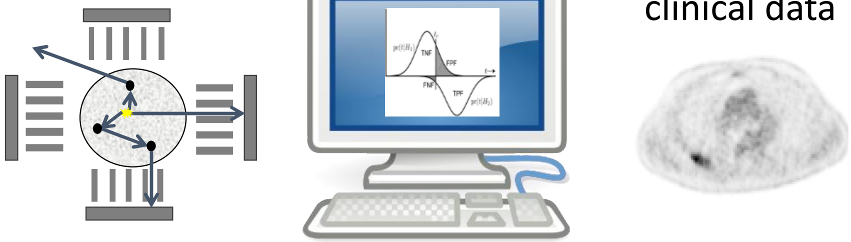
Transform nuclear medicine

The road ahead: Some important needs

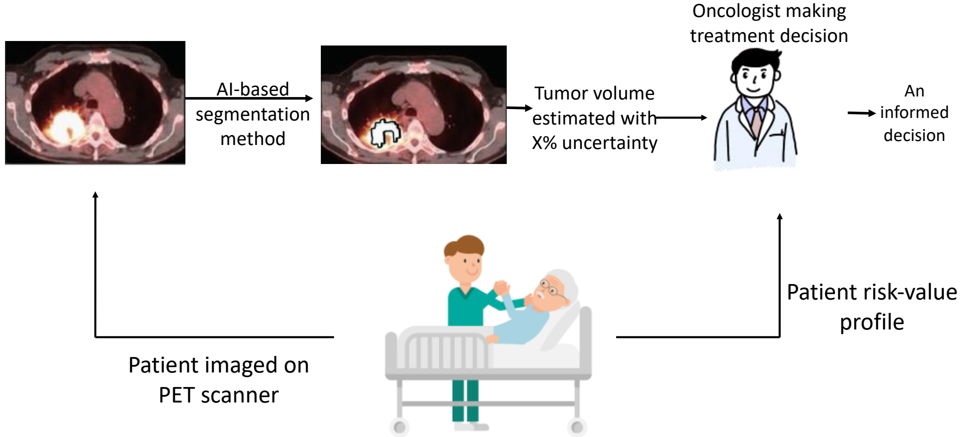
Simulation tools

Methods to extract task-specific information

Methods to evaluating with clinical data

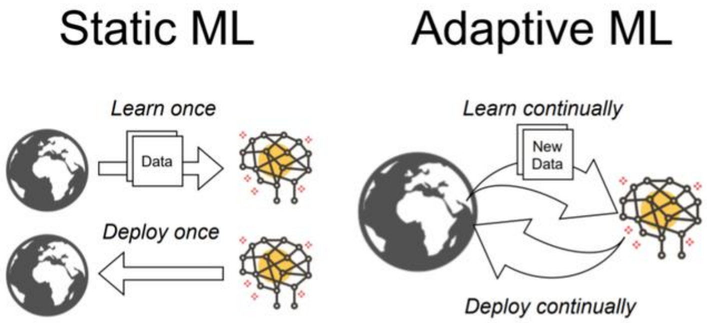


New methods for objective task-based evaluation



Methods to compute uncertainty of AI for clinical decision making

Inter-disciplinary collaborations and community-wide efforts for multi-center evaluations



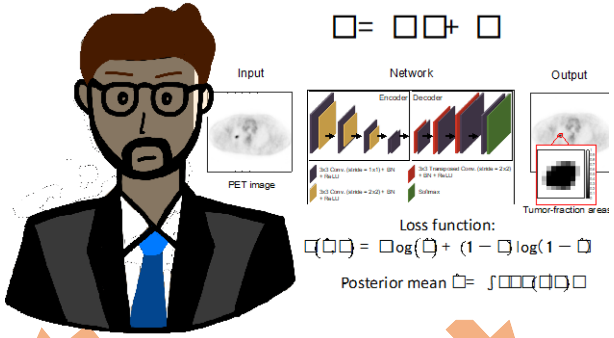
Develop evaluation strategies to adapt to the changing AI landscape

Wishlist of a computational nuclear-medicine (NM) imaging scientist interested in evaluation of AI algorithms

Access to list-mode data in a standardized format

Multi-disciplinary, multi-institutional collaborative opportunities

Support to develop technologies for evaluating AI for NM (virtual clinical trials, clinical and post-deployment evaluation)



Large, representative, diverse, high quality multi-center databases that model data and concept drifts.
A MIDRC for NM

Platforms for dissemination of evaluation technologies

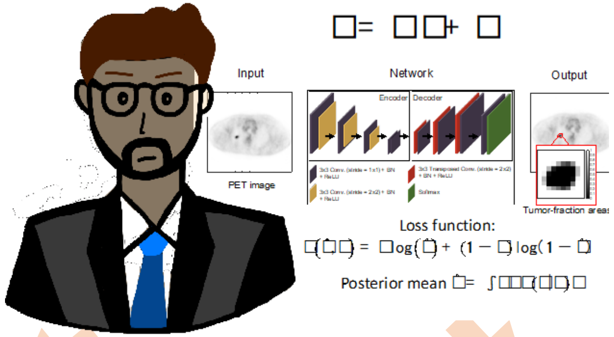
Support to develop methods for ethical evaluation of AI algorithms

Wishlist of a computational nuclear-medicine (NM) imaging scientist interested in evaluation of AI algorithms

Access to list-mode data in a standardized format

Multi-disciplinary, multi-institutional collaborative opportunities

Support to develop technologies for evaluating AI for NM (virtual clinical trials, clinical and post-deployment evaluation)



Large, representative, diverse, high quality multi-center databases that model data and concept drifts.
A MIDRC for NM

Platforms for dissemination of evaluation technologies

Support to develop methods for ethical evaluation of AI algorithms

Summary

- AI in nuclear medicine presents immense and exciting opportunities
- Rigorous evaluation is imperative for clinical translation of these opportunities
- Multiple ongoing efforts towards improving the evaluation of AI algorithms for nuclear medicine
- The road ahead provides a path to use AI for transforming nuclear medicine
- Exciting time to be in nuclear medicine!

Acknowledgements

- SNMMI AI Taskforce Evaluation Team
- SNMMI AI-Dosimetry Taskforce
- SNMMI Nuclear Medicine Data Standardization Initiative
- Attendees of SPIE Image Perception, Observer Performance, and Technology Assessment conference
- NIH NIBIB R01 EB031051, R56EB028287, R21 EB024647 (Trailblazer award)

SNMMI Value Initiative

SOCIETY OF NUCLEAR MEDICINE & MOLECULAR IMAGING

Thank you and welcome feedback and questions!

jhalab.wustl.edu

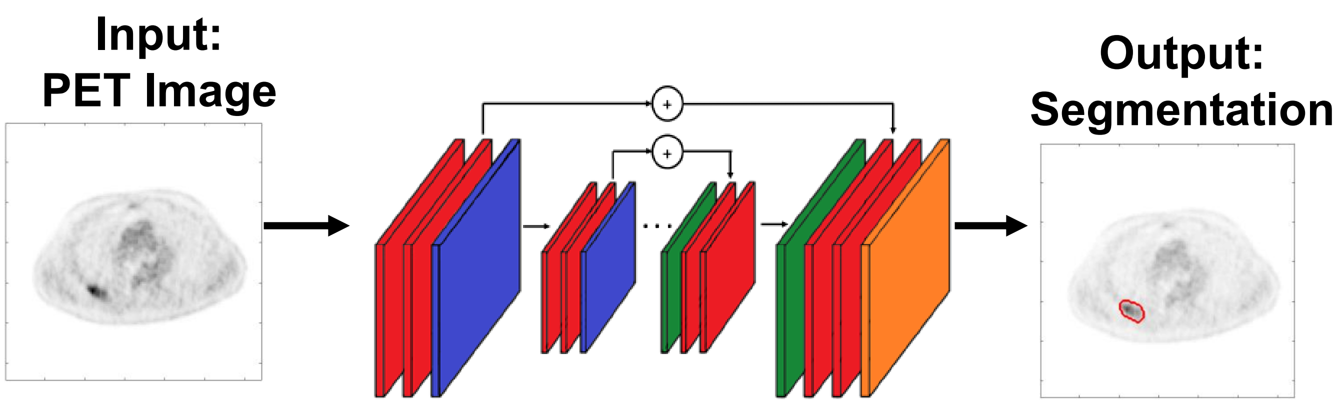
a.jha@wustl.edu

jhalab_washu

Backup slides

Class 1: Proof-of-concept evaluation

Objective: Demonstrate technical innovations of new methods



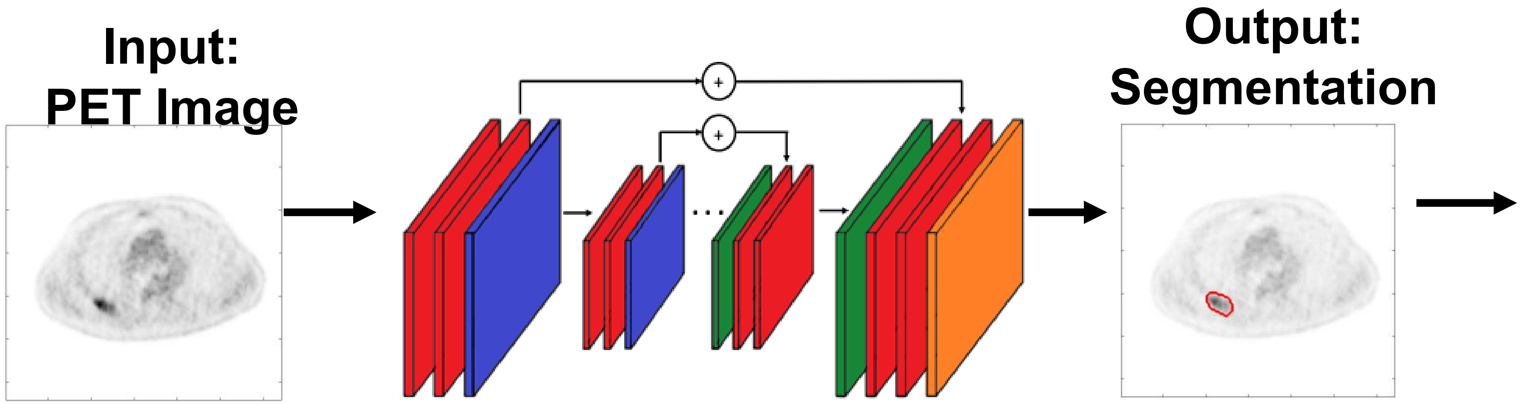
Proof of concept evaluation

Is this segmentation accurate?

Example claim: An AI-based PET segmentation method evaluated on patients with locally advanced lung cancer acquired on a single scanner with single-reader evaluation yielded mean dice scores of X (95% confidence intervals)

Class 2: Technical evaluation

Objective: Quantify technical factors such as accuracy, reproducibility and repeatability of the method on the specific clinical task



Metabolic tumor volume (MTV)

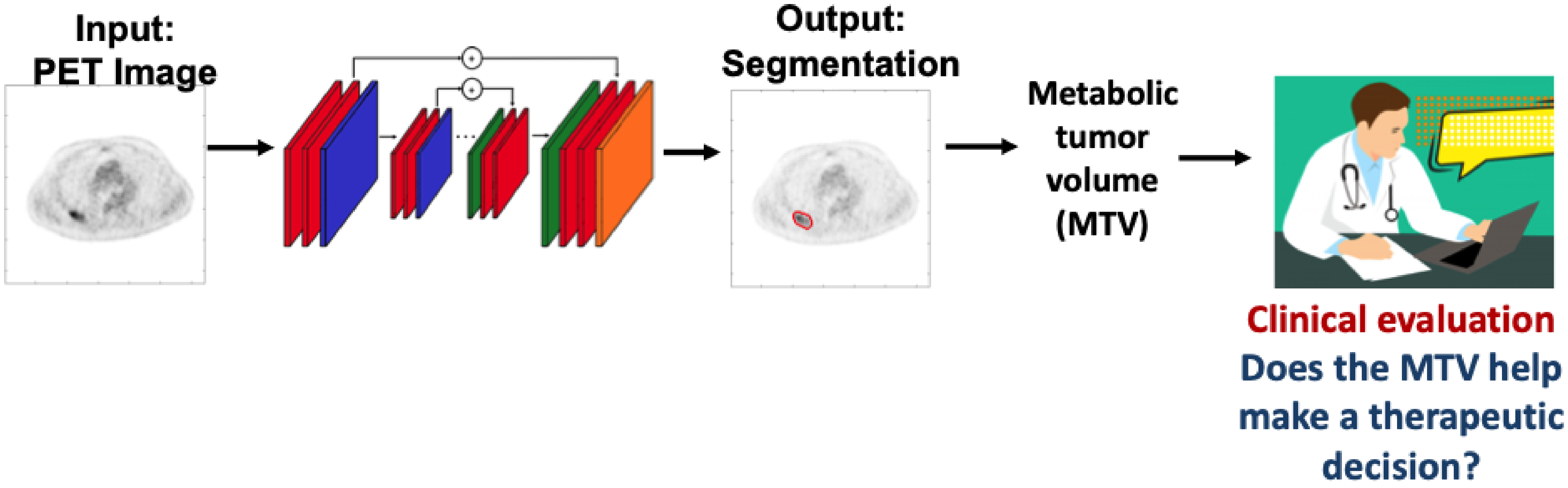
Technical evaluation

Is this MTV accurate, repeatable, reproducible?

Example claim: An AI-based PET segmentation method yielded MTV values with a normalized bias of X% (95% confidence intervals) as evaluated using an anthropomorphic thoracic physical phantom conducted on a single scanner in a single center

Class 3: Clinical evaluation

Objective: Quantify efficacy of the method for making clinical decisions

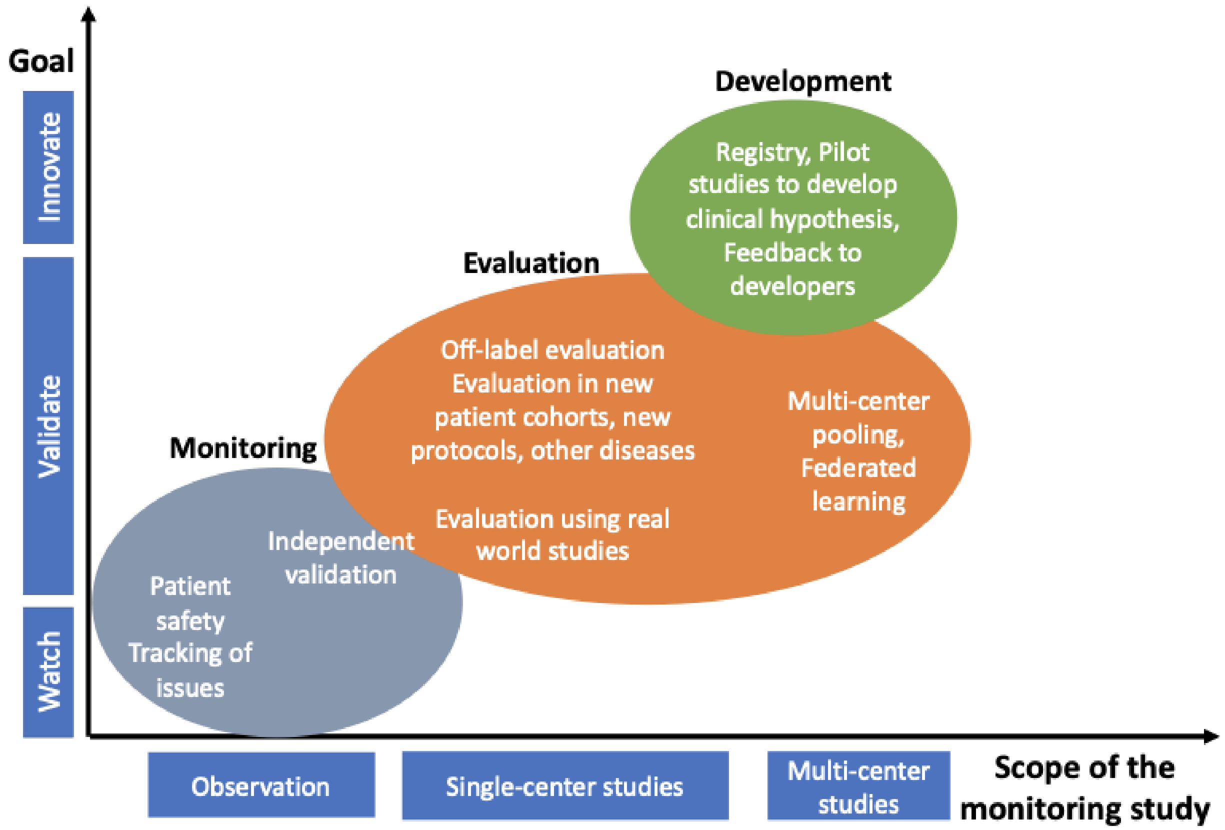


Example claim: Early change in MTV measured from FDG-PET images with an AI-based segmentation method yielded an increase in AUC from X to Y, with a change Δ (95% CIs of Δ) in predicting overall survival in patients with locally advanced lung cancer, as evaluated using a prospective observational study

Class 4: Post-deployment evaluation

Objectives

- Monitoring, detecting technical issues, potential bugs, reportable events, opportunities for improvement
- Evaluating off-label use
- Provide feedback for development



AI in Nuclear Medicine – An Academic's Perspective

Joyita Dutta, Ph.D.

Associate Professor

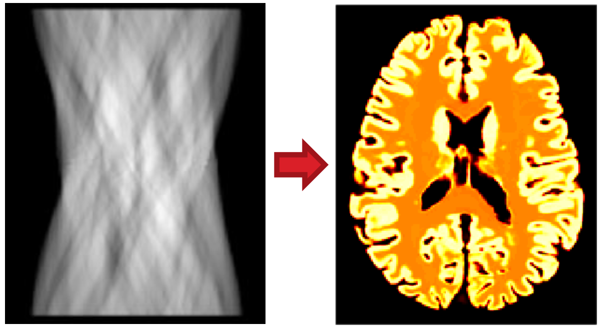
Biomedical Imaging and Data Science Lab (BIDSLab), Department of Electrical and Computer Engineering, University of Massachusetts Lowell

Gordon Center for Medical Imaging, Massachusetts General Hospital | Harvard Medical School

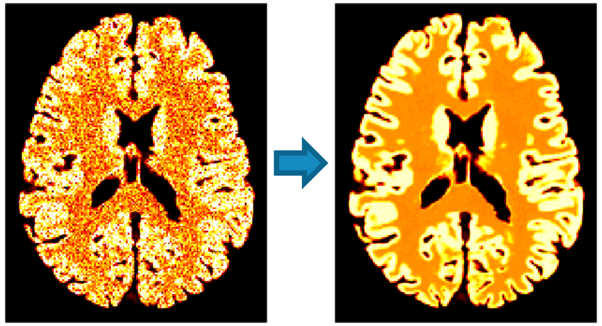
Common Inverse Problems in Medical Imaging

- Estimation/Regression

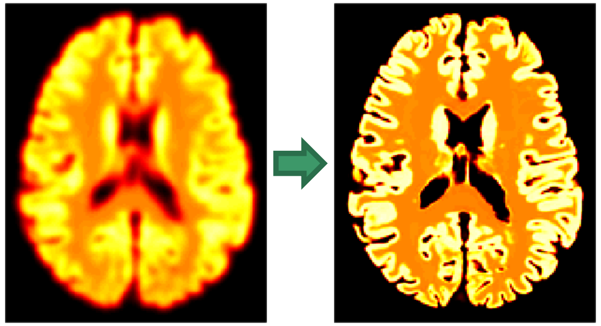
Reconstruction



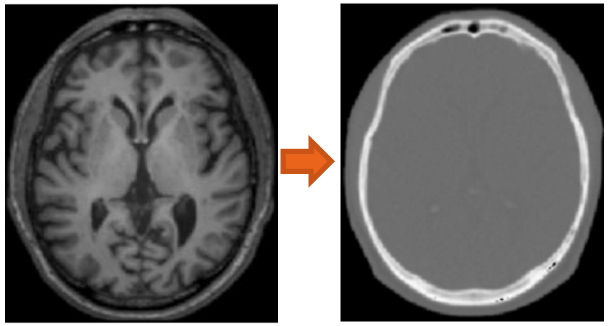
Denoising



Deblurring

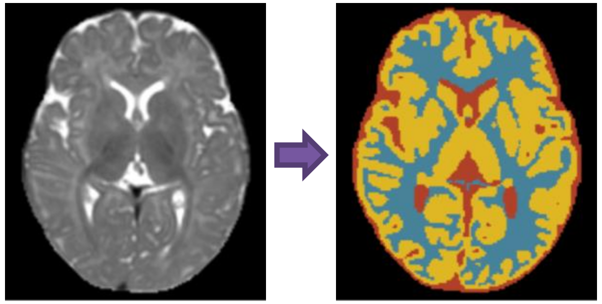


Synthesis

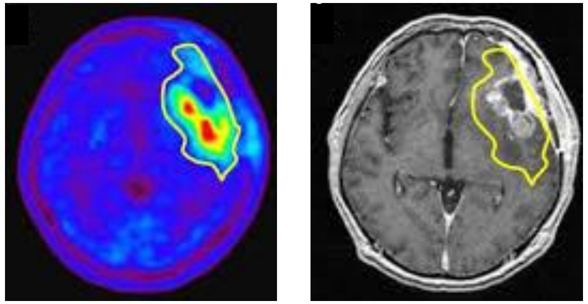


- Detection/Classification

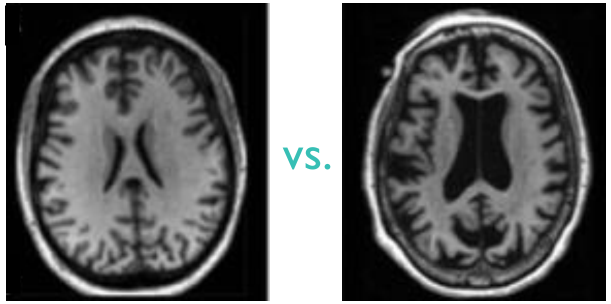
Segmentation



Lesion Detection

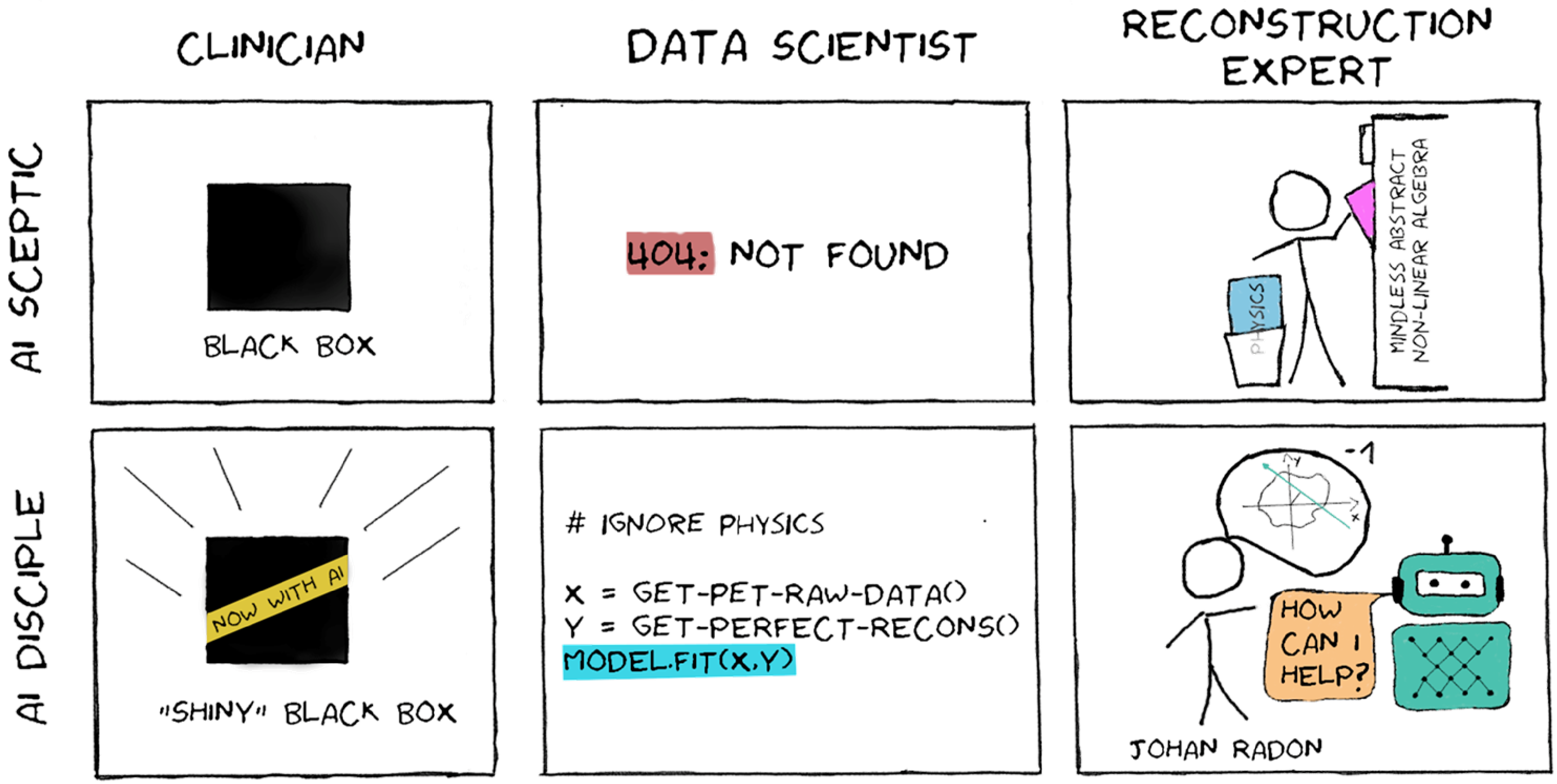


Disease Staging



Normal control　　Alzheimer's patient

Differing Perspectives



Source: https://figshare.com/articles/figure/AI_in_PET_image_reconstruction_v1_0/14685915

AI in Image Reconstruction, Denoising, and Deblurring

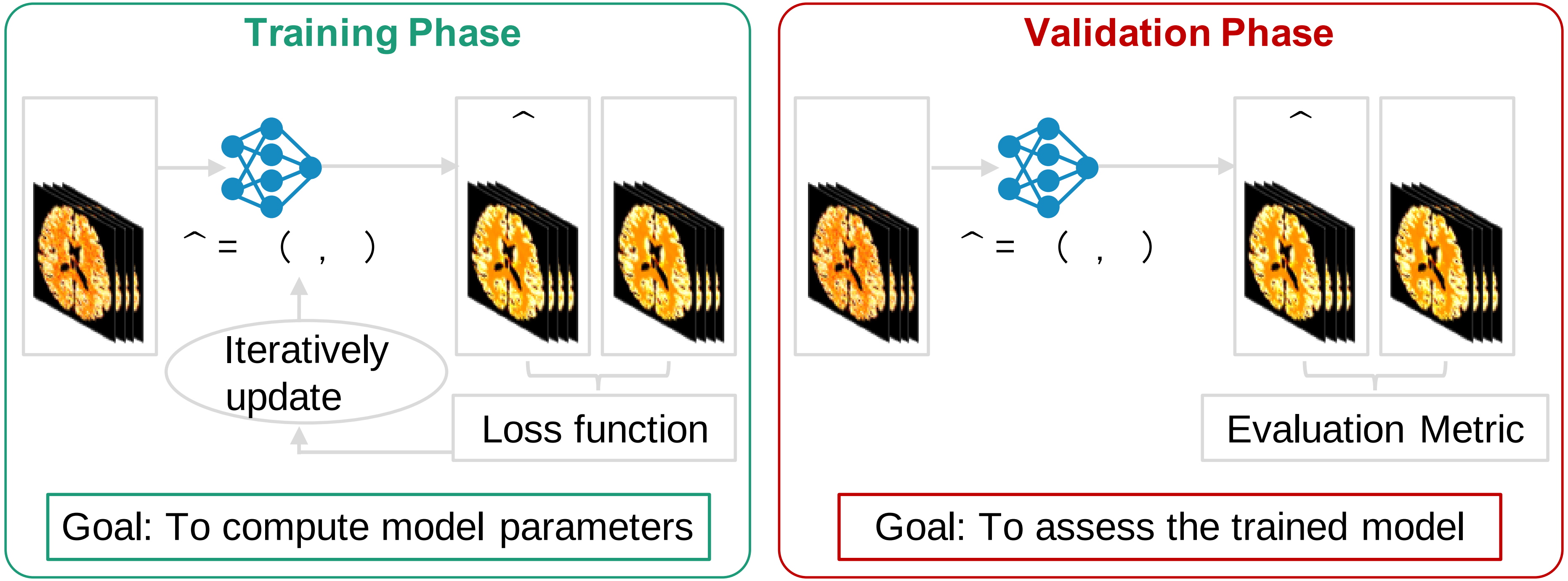
Reconstruction

- End-to-end reconstruction models: AUTOMAP, DeepPET etc.
- Physics-informed AI models: AI-based penalties/priors, unrolled networks, etc.

Denoising/Deblurring

- Supervised low-count to high-count mapping models
- Supervised low-resolution to high-resolution mapping models
- Unsupervised/self-supervised denoising and deblurring models

Active Areas of Research



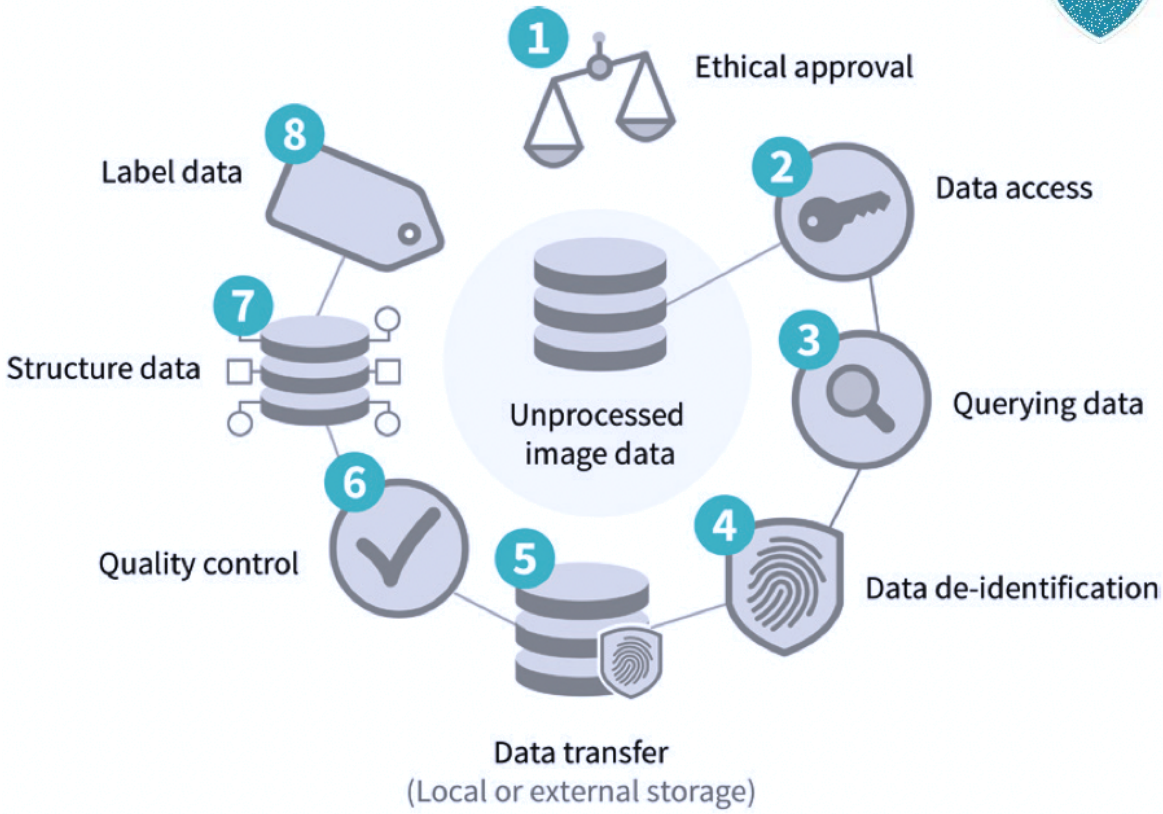
- New architectures, loss functions, evaluation metrics
- Adapting models to work for different resolution and noise levels and across scanners/site/cohorts
- Unsupervised alternatives: Masking techniques, DIP, Noise2Noise, etc.

Breakthroughs

- Harnessing cross-modality information
- Harnessing cross-tracer information
- New data acquisition paradigms with reduced duration or dose
- Speed*

Challenge I: Data Sharing

- Privacy
- Informed consent
- Data ownership



- Questions:
  - Who should control and profit from deidentified clinical data?
  - Can secondary use of clinical data be treated as a form of public good to be used for the benefit of future patients and not to be sold for profit or under exclusive arrangements?

Image source: Willemink et al., *Radiology*, 2020; 295:675–682

Toward Federated Learning

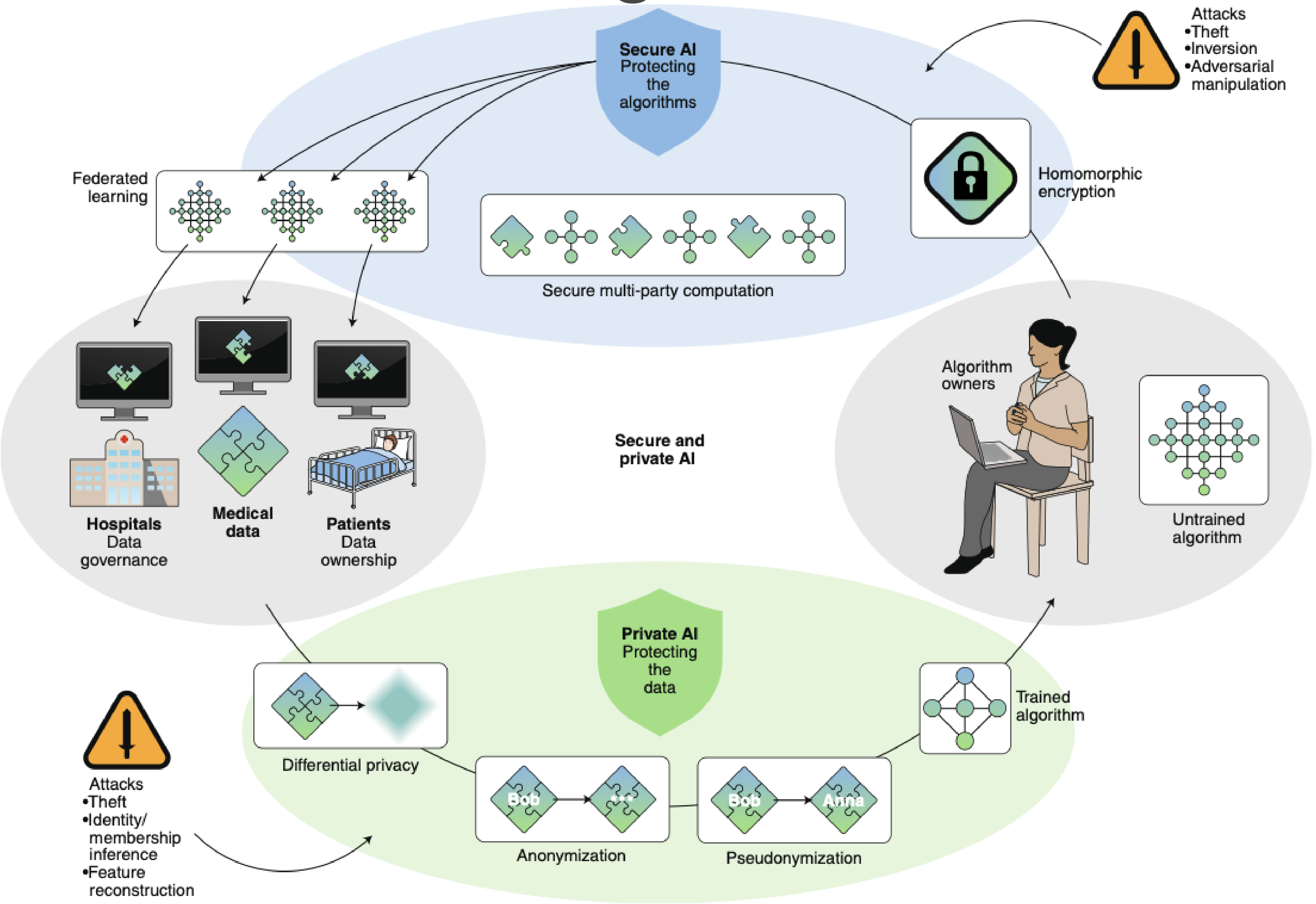


Image source: Kaissis et al., *Nature Machine Intelligence*, 2020; 2:305–311

Challenge II: Trustworthiness

- Need to ensure data volume, variety, and veracity
- Need for reproducibility
- Need for generalizability
- Bias control
- Transparency
- Explainability

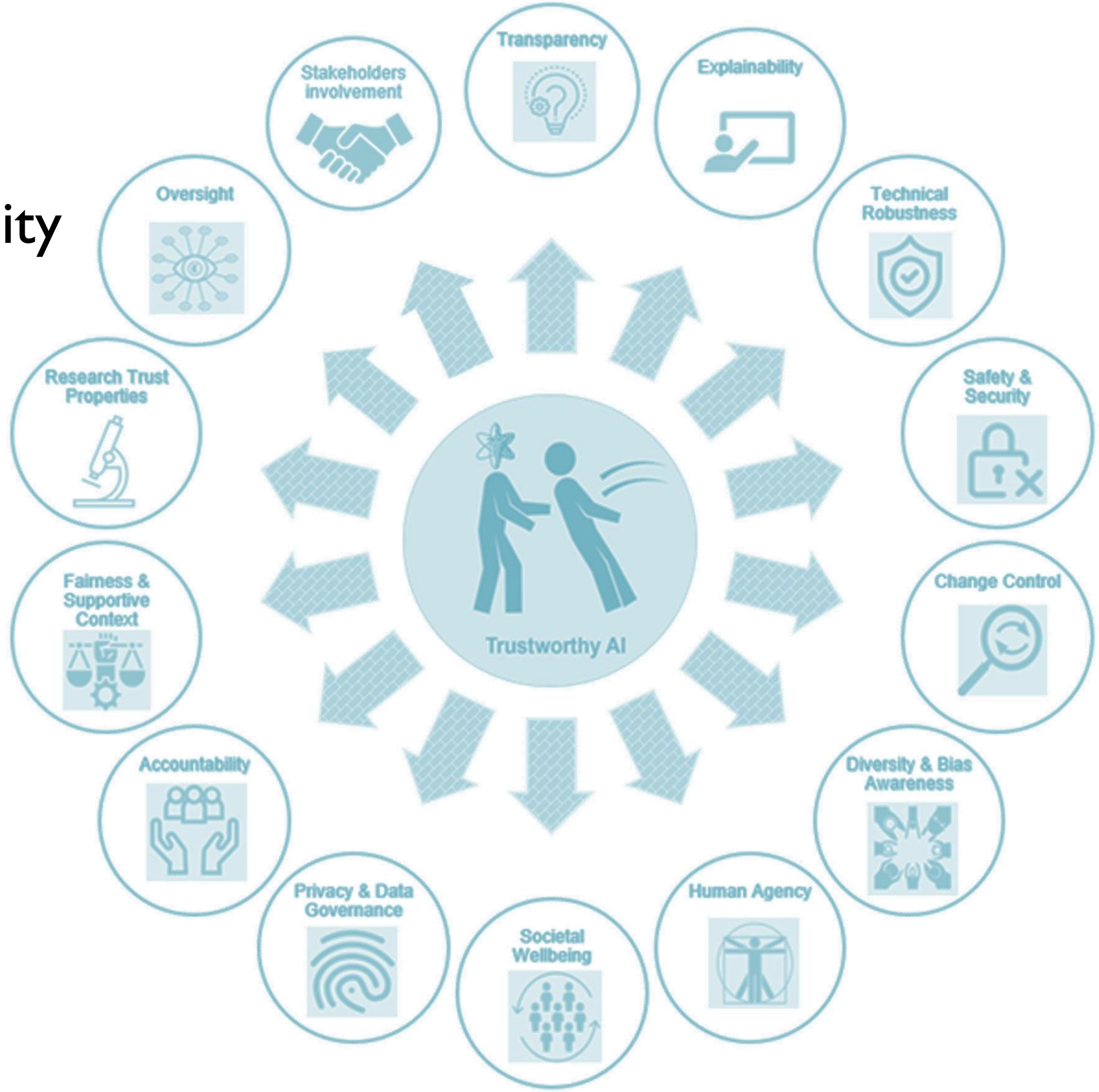


Image source: Jones et al., *PET Clinics*, 2022; 17:P1–12

Challenge III: Standardization

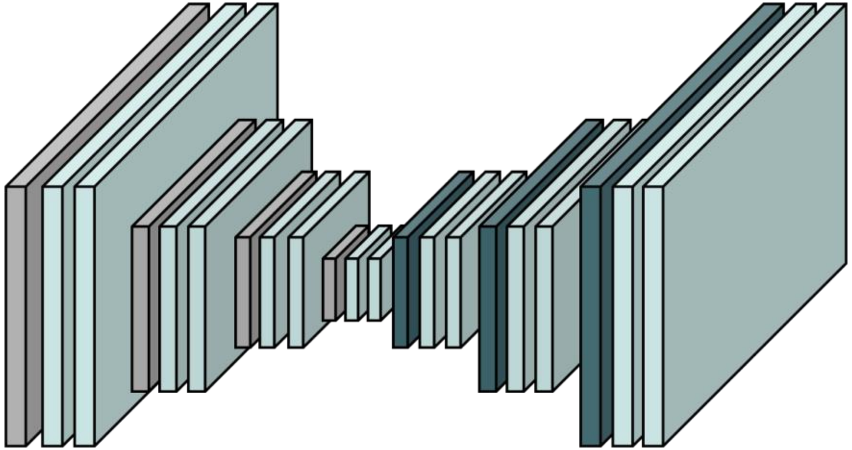
- Need for universal benchmarking standards in nuclear medicine
  - De-identified datasets
  - Agreed upon evaluation metrics
  - Independent secondary validation requirements

Emerging Areas and Future Directions

- Unsupervised, weakly-supervised, and self-supervised learning models and network architectures
- Transfer learning paradigms
- Federated learning protocols
- Interpretable machine learning models

Pitfalls in Developing Artificial Intelligence Algorithms in Nuclear Medicine

SNMMI Artificial Intelligence Summit
March 21, 2022

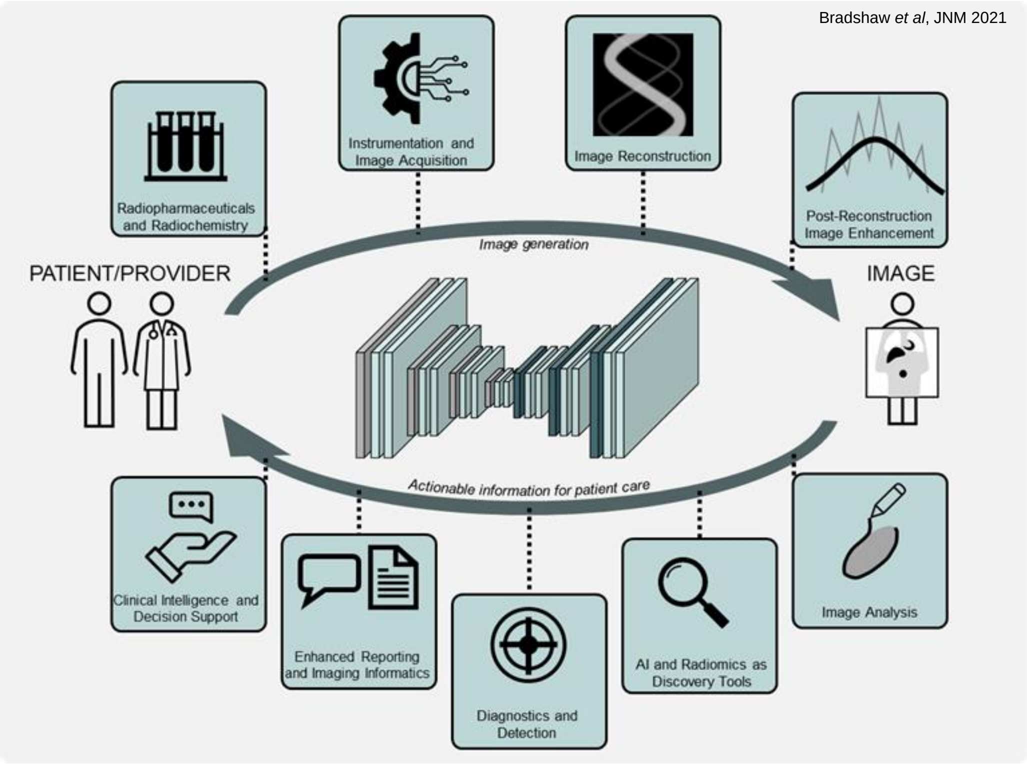


Tyler Bradshaw, PhD, DABSNM
Scientist II, Imaging Sciences
Department of Radiology
University of Wisconsin

Disclosures

Research support from GE Healthcare

The promise of AI in nuclear medicine

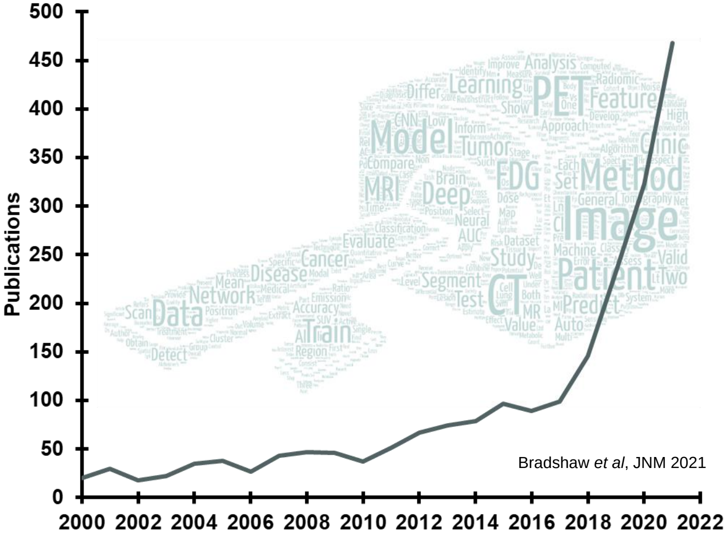


The problem with AI

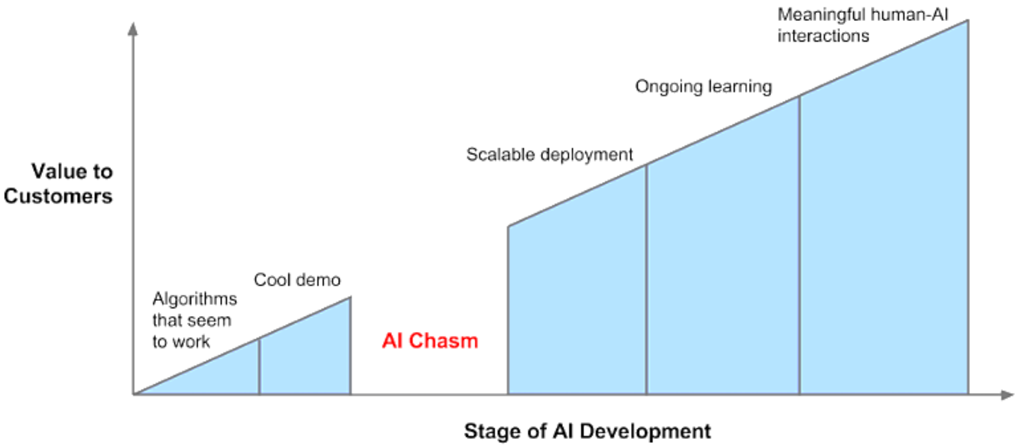
Numerous models have shown promise

Few models are ultimately useful

Publications on AI in NM



AI Chasm



https://techcrunch.com/2017/01/05/crossing-the-ai-chasm/

Why do many models not work?

1. Poor reproducibility

I run your codes on your data and do not come to the same conclusion

? Definitions are inconsistent

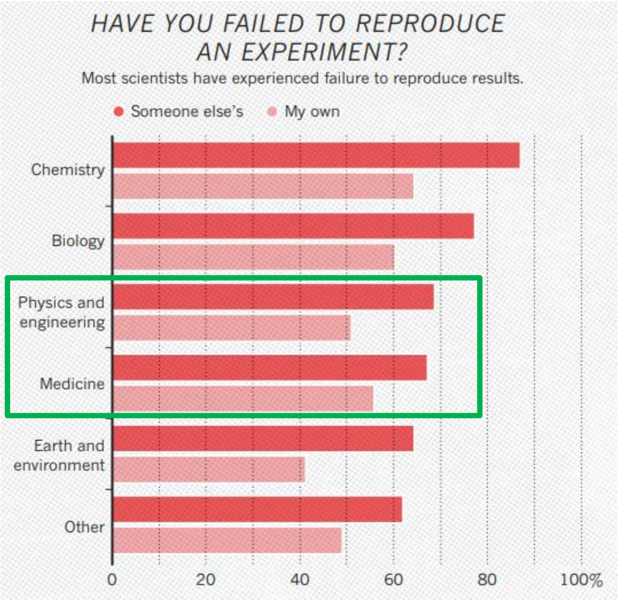
2. Poor replicability

I run the same methods on similar data and do not come to the same conclusion

3. Poor generalizability

I run your model on a different population and do not come to the same conclusion

It is a serious problem in many fields

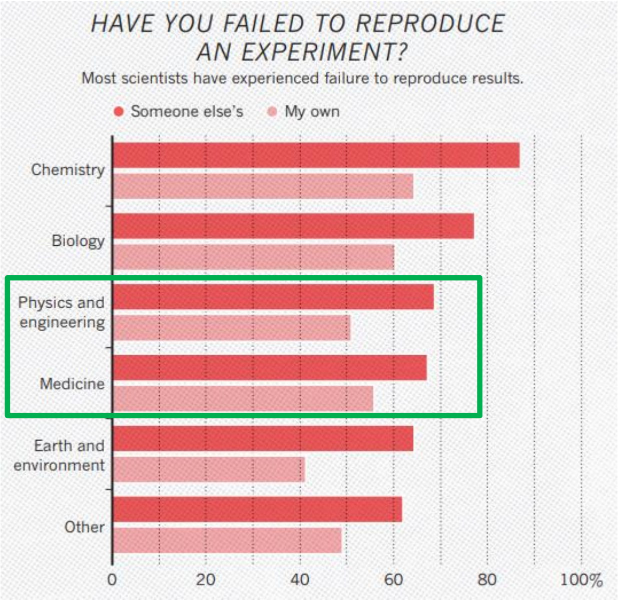


Baker M, Nature; 533:452, 2016

Reproducibility

Same code, same data

It is a serious problem in many fields



Baker M, Nature; 533:452, 2016

Science *Artificial intelligence faces reproducibility crisis*
Unpublished code and sensitivity to training conditions make many claims hard to verify

It's actively being addressed in CS

ML Reproducibility Challenge 2021

…inviting members of the community at large to select a paper and verify the empirical results and claims in the paper by reproducing the computational experiments, either via a new implementation or using code/data or other information provided by the authors.

RESCIENCE C

Reproducible Science is good. Replicated Science is better.

Reproducibility Challenge @ NeurIPS 2019
The Annual Machine Learning Reproducibility Challenge

The Machine Learning Reproducibility Checklist (v2.0, Apr.7 2020)

…what about in radiology???

Replicability

Same methods, similar data

Researcher tried to replicate 255 ML papers

– **64% replication success rate**

A Step Toward Quantifying Independently Reproducible Machine Learning Research

Edward Raff
Booz Allen Hamilton
raff_edward@bah.com
University of Maryland, Baltimore County
raff.edward@umbc.edu

Abstract

What makes a paper independently reproducible? Debates on reproducibility center around intuition or assumptions but lack empirical results. Our field focuses on releasing code, which is important, but is not sufficient for determining reproducibility. We take the first step toward a quantifiable answer by manually attempting to implement 255 papers published from 1984 until 2017, recording features of each paper, and performing statistical analysis of the results. For each paper, we did not look at the authors code, if released, in order to prevent bias toward discrepancies between code and paper.

Generalizability

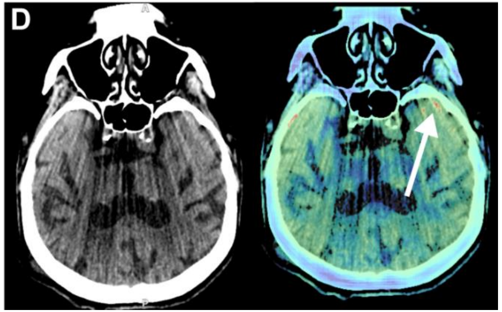
Same model, new population

"Dataset shift"

Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Intracranial Hemorrhage

Andrew F. Voter, PhD[a], Ece Meram, MD[b], John W. Garrett, PhD[b], John-Paul J. Yu, MD, PhD[b,c,d]

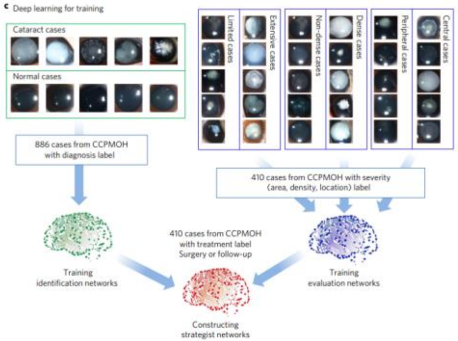
- 15x more FPs than a physician
- 3x higher error in patients with history of surgery
- 81% PPV vs 92% from previous studies



Diagnostic Efficacy and Therapeutic Decision-making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: A Multicentre Randomized Controlled Trial

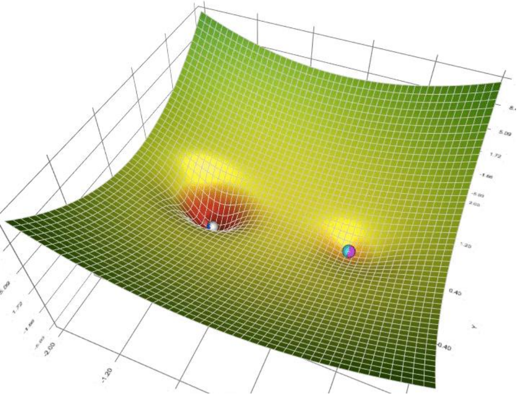
Haotian Lin[a,1], Ruiyang Li[a,1], Zhenzhen Liu[a,1], Jingjing Chen[a,1], Yahan Yang[a,1], Hui Chen[a,1], Zhuoling Lin[a], Weiyi Lai[a], Erping Long[a], Xiaohang Wu[a], Duoru Lin[a], Yi Zhu[a,b], Chuan Chen[a,b], Dongxuan Wu[c], Tongyong Yu[c], Qianzhong Cao[a], Xiaoyan Li[a], Jing Li[a] ... Yizhi Liu[a]

- Randomized controlled trial
- Accuracy: expected 99%, got 87%

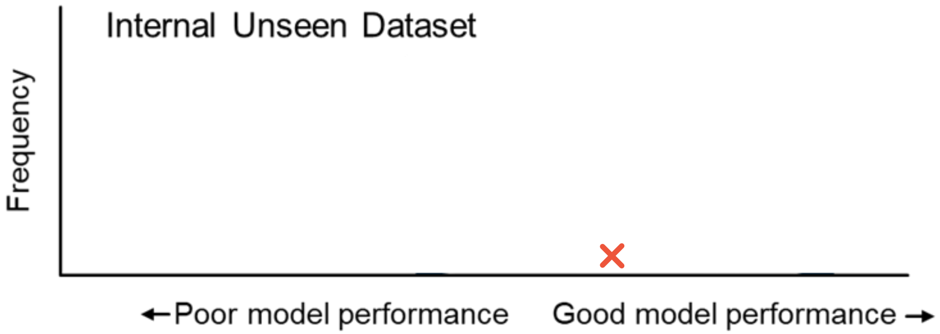


Why is this happening?

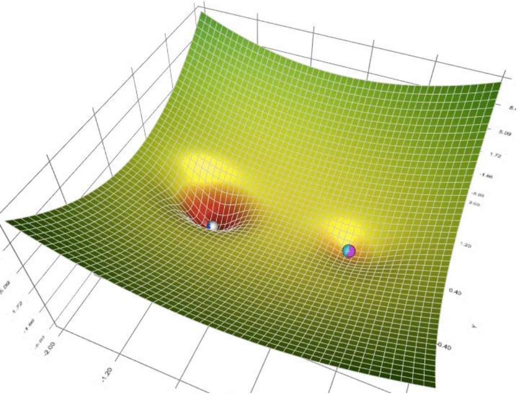
Why is this happening?



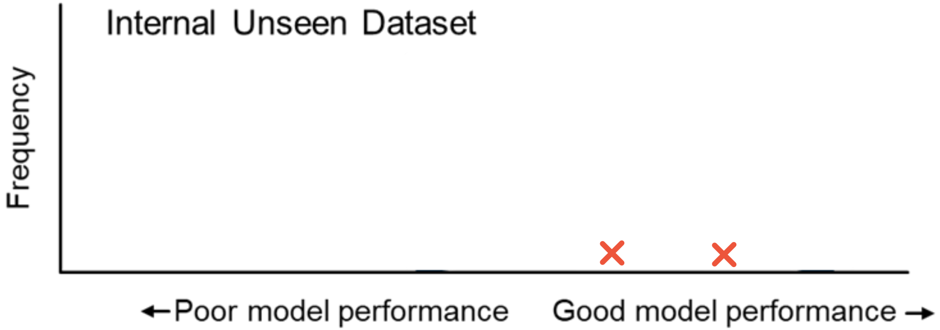
https://towardsdatascience.com/a-visual-explanation-of-gradient-descent-methods-momentum-adagrad-rmsprop-adam-f898b102325c



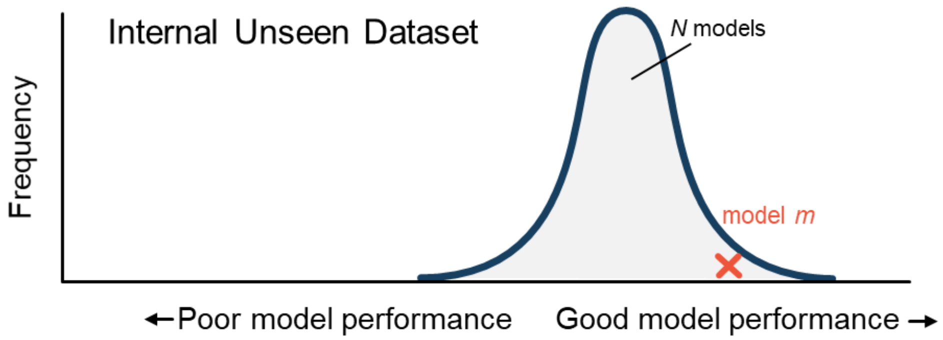
Why is this happening?



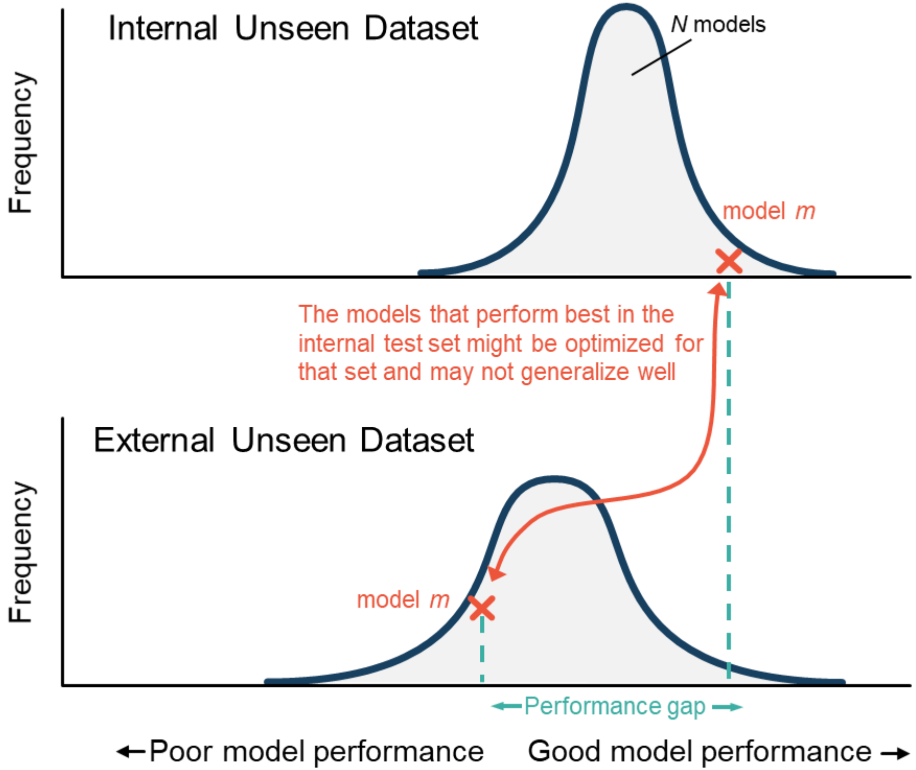
https://towardsdatascience.com/a-visual-explanation-of-gradient-descent-methods-momentum-adagrad-rmsprop-adam-f898b102325c



Why is this happening?



Why is this happening?



What can we do about it?

THE STATE OF THE ART

Nuclear Medicine and Artificial Intelligence: Best Practices for Algorithm Development

Tyler J. Bradshaw[1], Ronald Boellaard[2], Joyita Dutta[3], Abhinav K. Jha[4], Paul Jacobs[5], Quanzheng Li[6], Chi Liu[7], Arkadiusz Sitek[8], Babak Saboury[9], Peter J.H. Scott[10], Piotr J. Slomka[11], John J. Sunderland[12], Richard L. Wahl[13], Fereshteh Yousefirizi[14], Sven Zuehlsdorff[15], Arman Rahmim[16], and Irène Buvat[17]

Table 2. Summary of recommendations.

Category	Topic	Recommendation
Study Design	Task definition	Collaborate with domain experts, stakeholders
	Study types	Publications should identify as development studies or evaluation studies
	Risk assessment	A study's degree of rigor should depend on the risk the algorithm poses to patients
	Statistical plan	Prospective studies should preregister statistical analysis plans
Data Collection	Bias anticipation	Collect data belonging to classes/groups that are vulnerable to bias
	Training set size estimation	Based on trial and error, or prior similar studies
	Evaluation set size estimation*	Guided by statistical power analysis
	Data decisions	Inclusion/exclusion criteria should be justified, objective, and documented
Data Labeling	Reference standard	Labels should be regarded as sufficient standards of reference by the field
	Label quality	Label quality should be justified by the application, study type, and clinical claim (Figure 4)
	Labeling guide*	Reader studies should produce a detailed guide for labelers
	Quantity/quality tradeoff	Consider multiple labelers (quality) over greater numbers (quantity)
Model Design	Model comparison*	Development studies should explore and compare different models
	Baseline comparison	Complex models should be compared with simpler models and/or standard-of-care
	Model selection	The model selection and hyperparameter tuning techniques should be reported
	Model stability	Repeated training with random initialization is recommended
	Ablation study*	Development studies focusing on novel architectures should perform ablation studies
Model Training	Cross validation*	Cross validation should be used for development studies; preserve data distribution across splits
	Data leakage	Information leaks from the test/evaluation set to the model during training must be avoided
Model Testing and Interpretability	Test set	Should have same data/class distribution as the target population; high quality labels
	Target population	The target population should be explicitly defined
	External sets	External sets are recommended for evaluating model sensitivity to dataset shift
	Evaluation metric	May consist of multiple metrics; often requires visual inspection of model output
	Model interpretability*	Interpretability is needed for clinical tasks
Reporting and Dissemination	Reporting	Follow published reporting guidelines/checklists
	Sharing*	Development studies must make code and models accessible
	Transparency	Be forthcoming about failure modes and population characteristics in training/evaluation sets
	Reproducibility checks	Journals should ensure that submitted materials are sufficient for replication

Final thoughts

Summary

- Poorly developed/validated algorithms can cause <u>distrust</u> and fear (users and patients)
- Overfitting + randomness makes deep learning <u>susceptible to low reproducibility, replicability, and generalizability</u>
- <u>Guidelines are needed</u> for best practices along all stages of development

Panel Questions

- How do we “raise the bar” for research quality without stifling innovation?
- How do we incentivize researchers to share data, codes, and models (i.e., a culture of sharing)?

Acknowledgements

SNMMI Artificial Intelligence Task Force, Team 2

- Abhinav Jha
- Arman Rahmim
- Arek Sitek
- Babak Saboury
- Bonnie Clark
- Chi Liu
- Irène Buvat
- Joyita Dutta
- John Sunderland
- Paul Jacobs
- Peter Scott
- Piotr Slomka
- Quanzheng Li
- Rich Wahl
- Ronald Boellaard
- Sven Zuehlsdorff

tbradshaw@wisc.edu

@TyBradshaw11

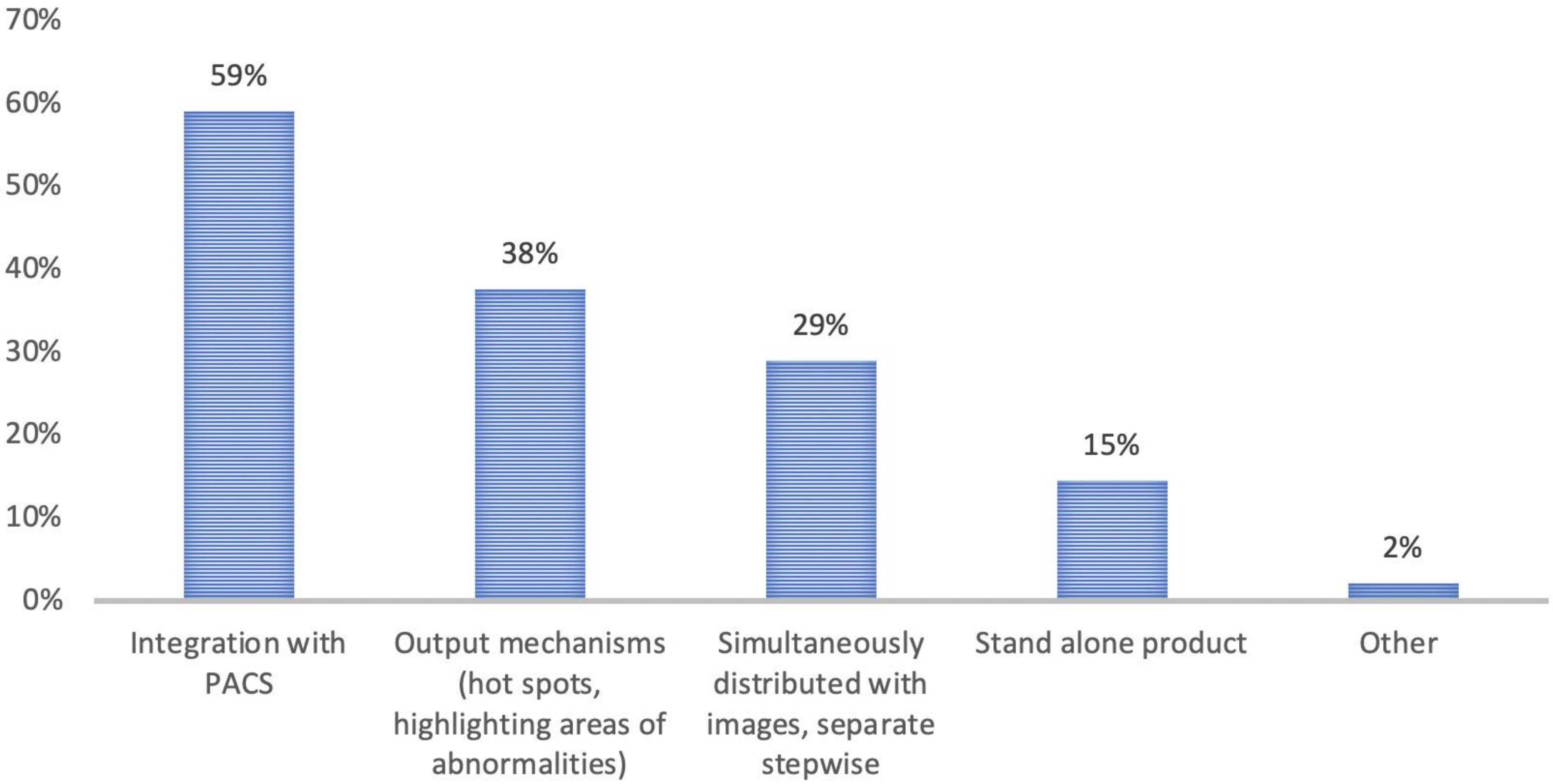
Translations of Artificial Intelligence-Based in Imaging Technologies

Chi Liu, PhD
Associate Professor
Radiology and Biomedical Imaging
Yale University

Questions to Panel 2 (Industry Representatives)

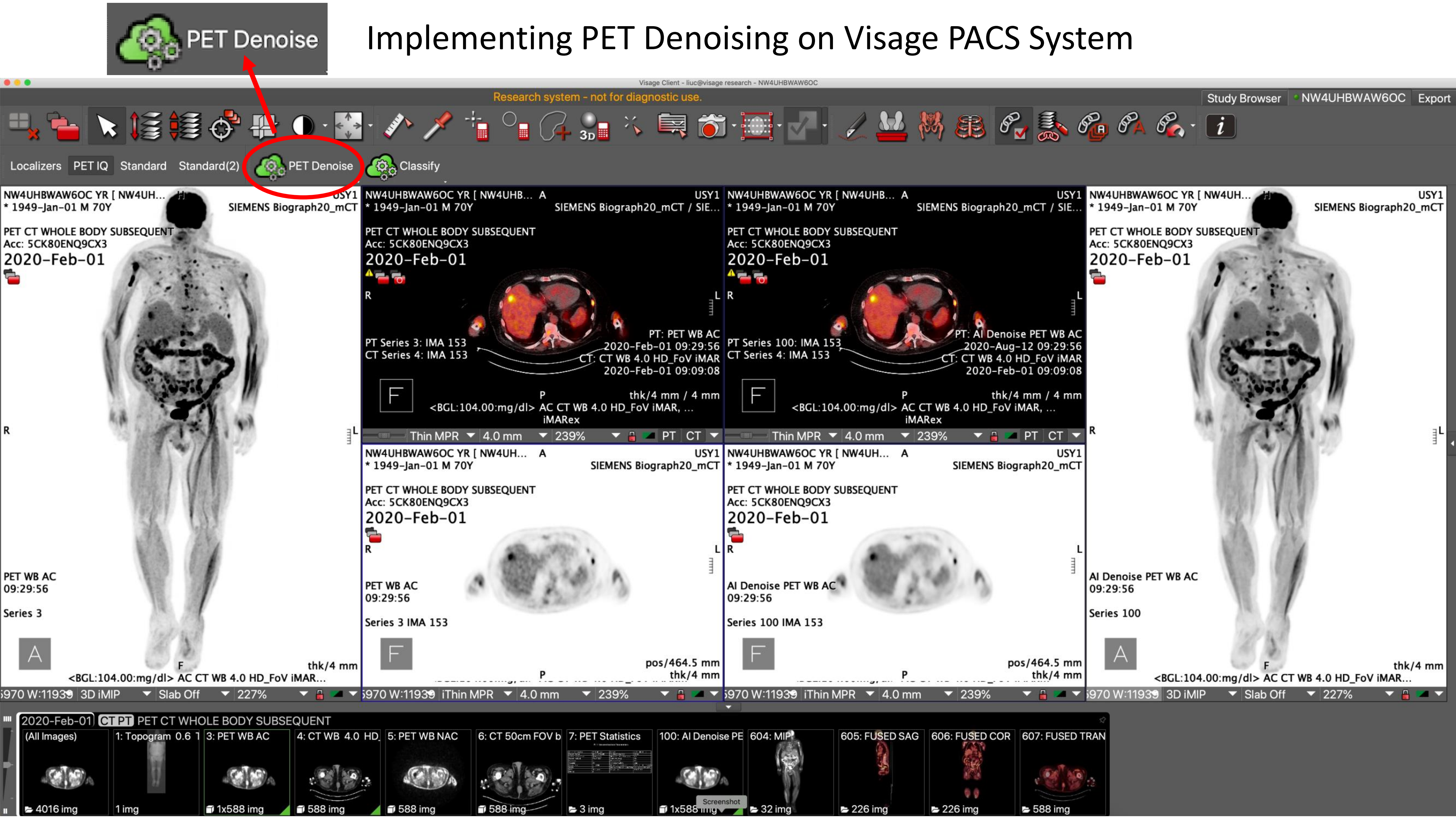
- Where to implement?
  - Scanner consoles
  - Workstations
  - Cloud
  - PACS
  - ...
- What data are accessible on the platform?
  - Images
  - Sinograms/Raw data
  - Listmode
  - Motion tracking signals?

AI'S IMPLEMENTATION IN PRACTICE



SNMMI AI Taskforce Survey, 186 responses

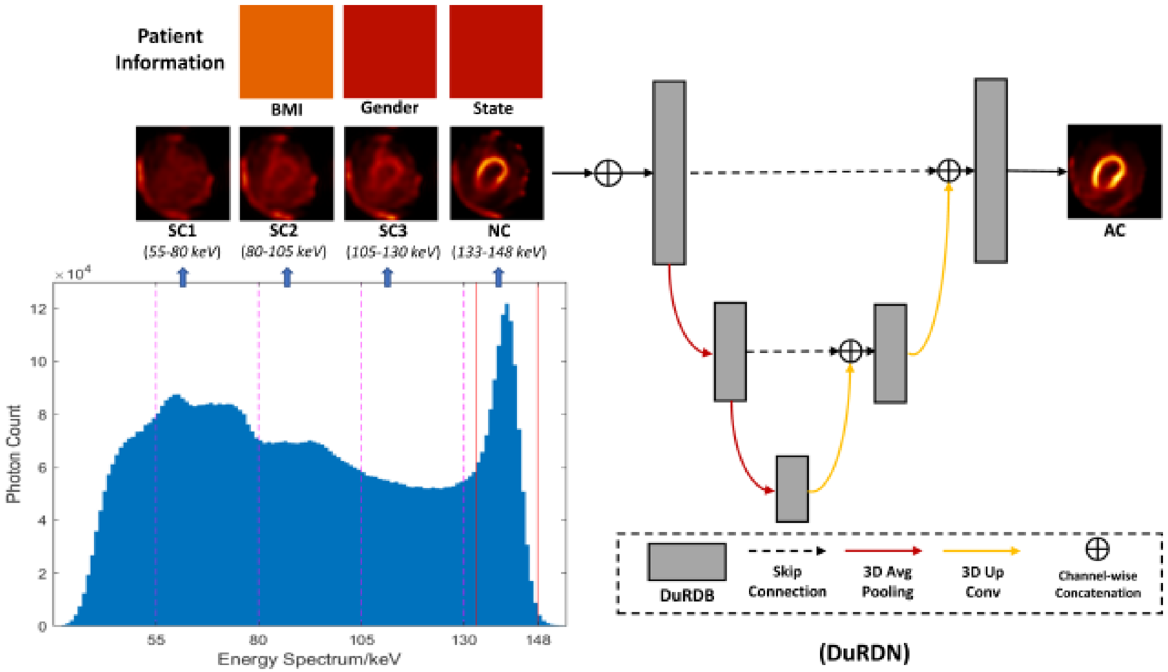
Implementing PET Denoising on Visage PACS System



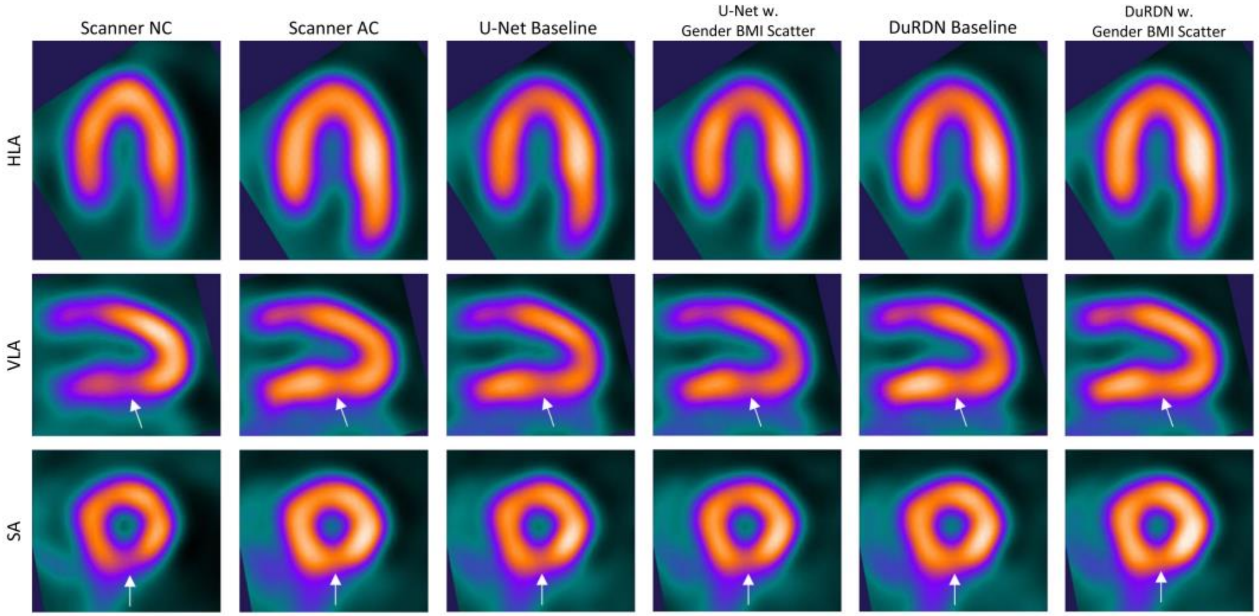
Direct prediction of attenuation-correction SPECT w/ scatter and non-imaging information

- *From ^{99m}Tc-tetrofosmin cardiac SPECT$_{NAC}$ MPI to SPECT$_{AC}$, for dedicated cardiac SPECT scanners*

Advanced algorithms, additional patient information incorporated



The schematic of our proposed AC workflow.



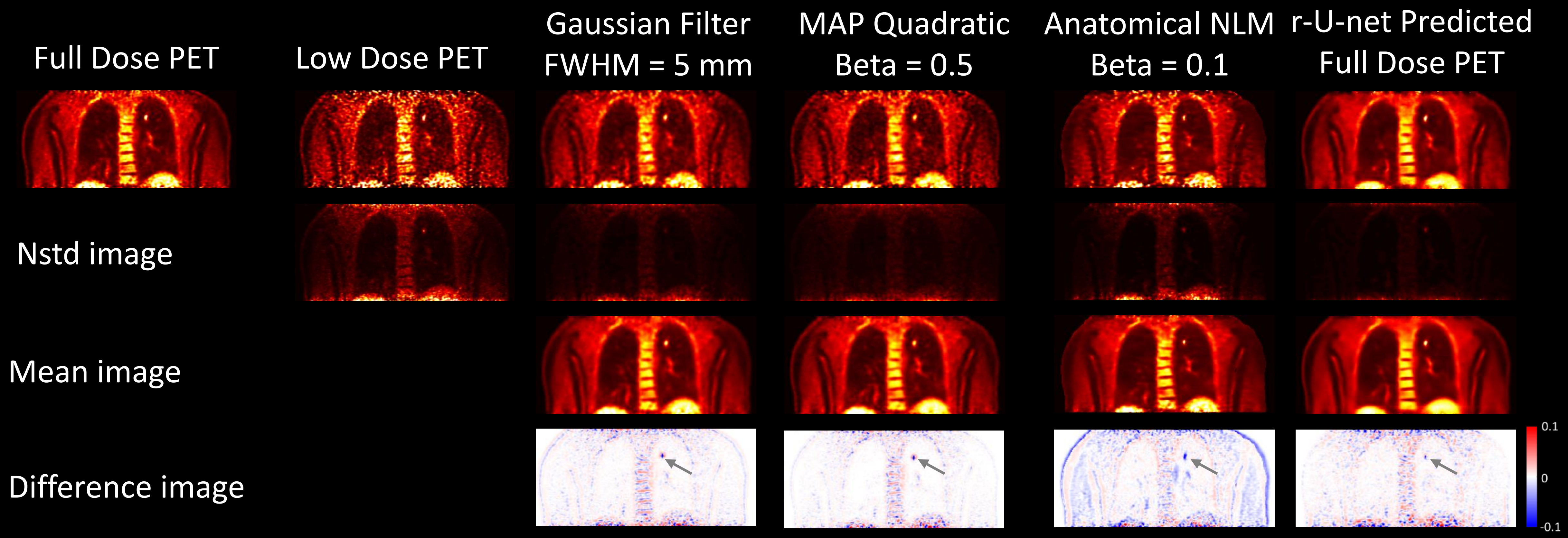
Visualization of NAC and AC SPECT images and polar maps.

Chen, Xiongchao, et al. "CT-free attenuation correction for dedicated cardiac SPECT using a 3D dual squeeze-and-excitation residual dense network." Journal of Nuclear Cardiology (2021): 1-16.

Questions to Panel 3: End users of AI (Physicians, technologists, hospital administrators)

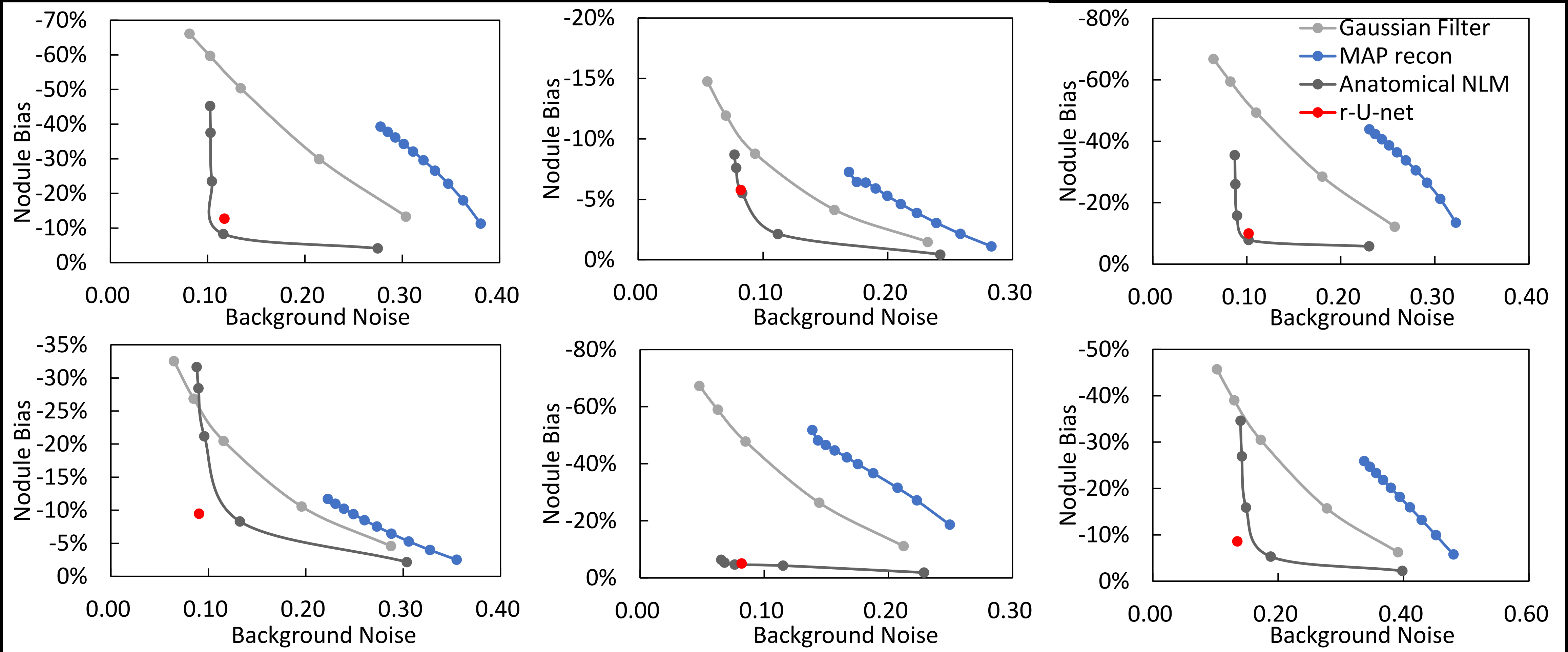
- How much improvement in image quality can impact clinical practice?
- If AI introduces bias, how much is acceptable?
- If AI introduces artifact, how much is acceptable?

Comparing Denoising Methods



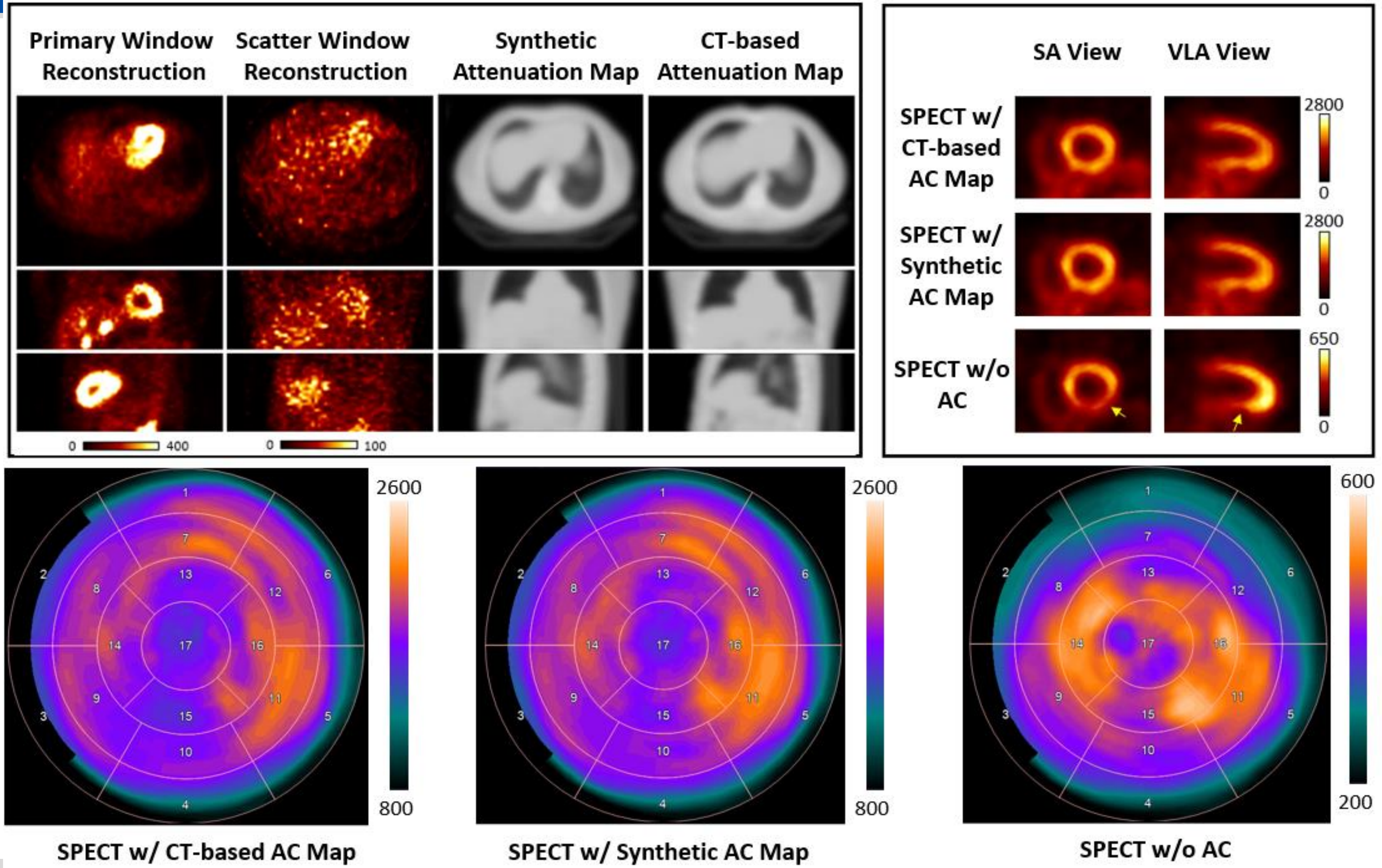
Wenzhuo Lu, et al. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. 2019 Phy. Med. Bio. 64 165019

Comparison with Existing Denoising Methods



Wenzhuo Lu, et al. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. 2019 Phy. Med. Bio. 64 165019

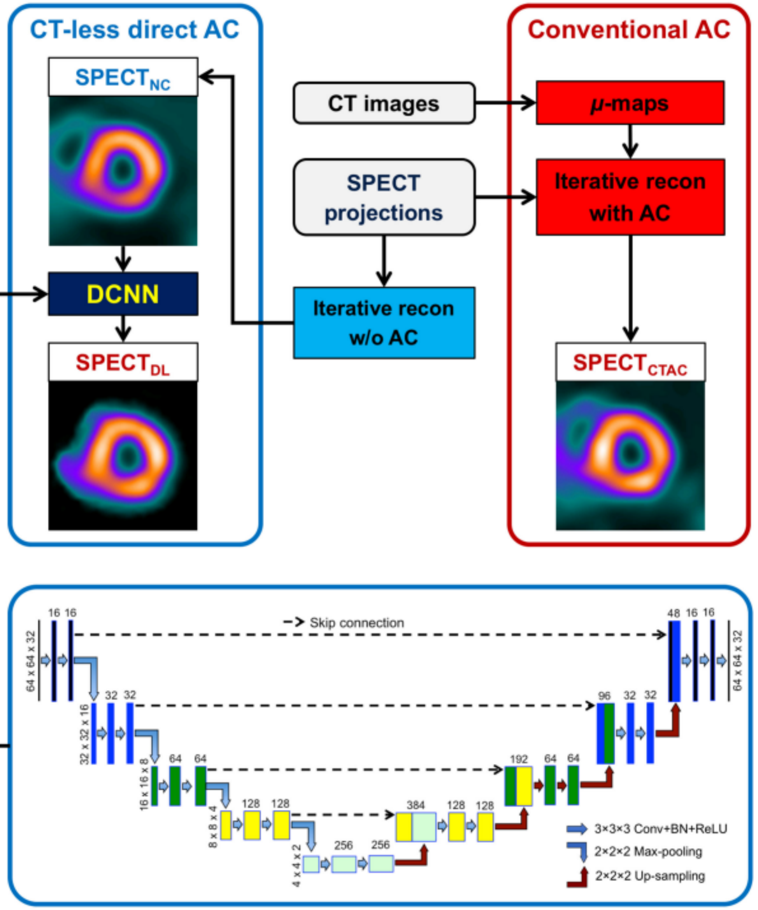
CT-less SPECT Attenuation Map Generation



L. Shi, et al. EJNMMI, Vol. 47 Issue 10, p2383-2395, 2020

Direct Prediction of Attenuation-Corrected Cardiac SPECT from Uncorrected SPECT

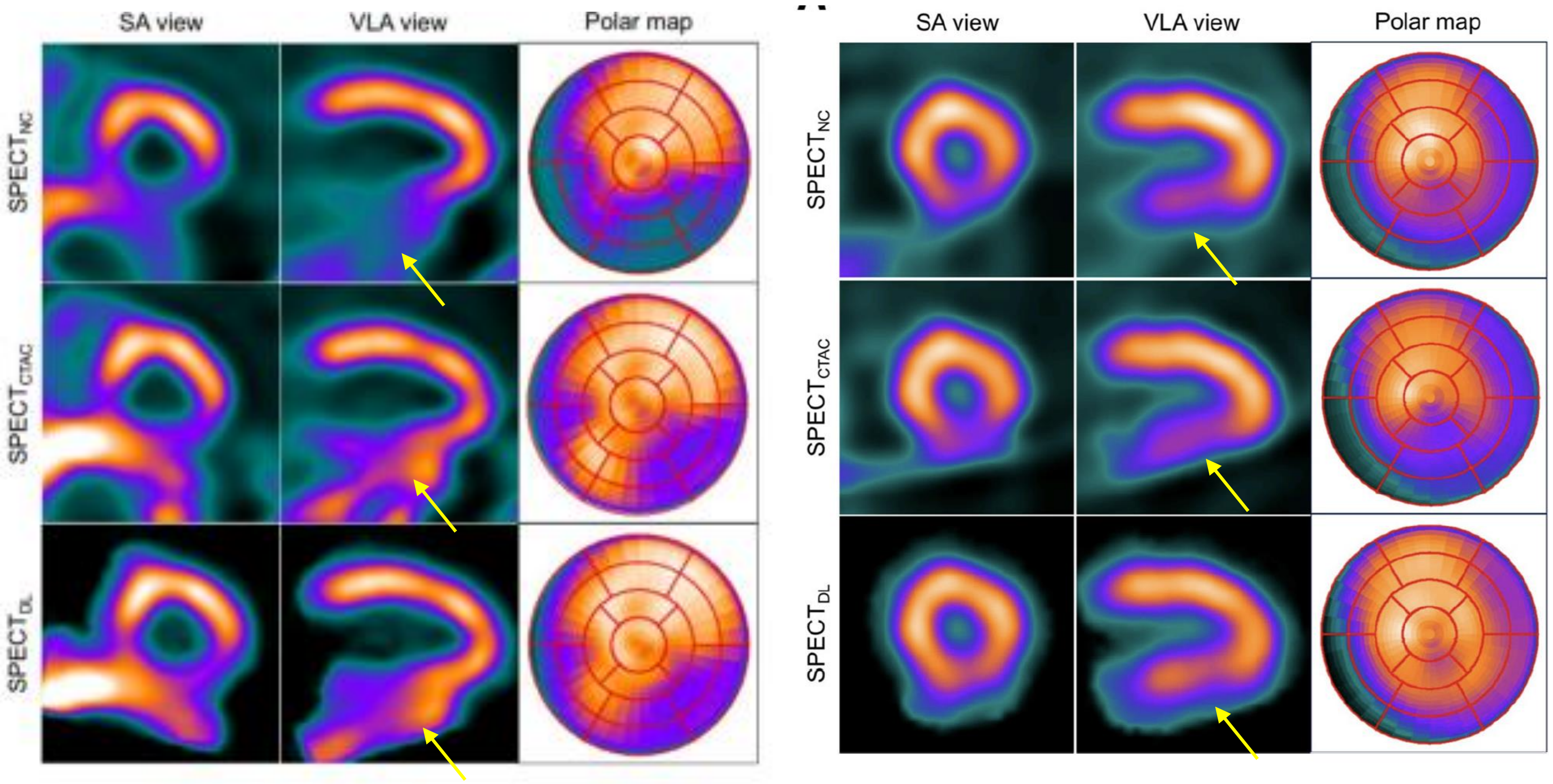
- From ^{99m}Tc-tetrofosmin cardiac $SPECT_{NAC}$ MPI to $SPECT_{AC}$, for dedicated cardiac SPECT scanners



Schematic of U-Net-based AC workflow.

Good Correction

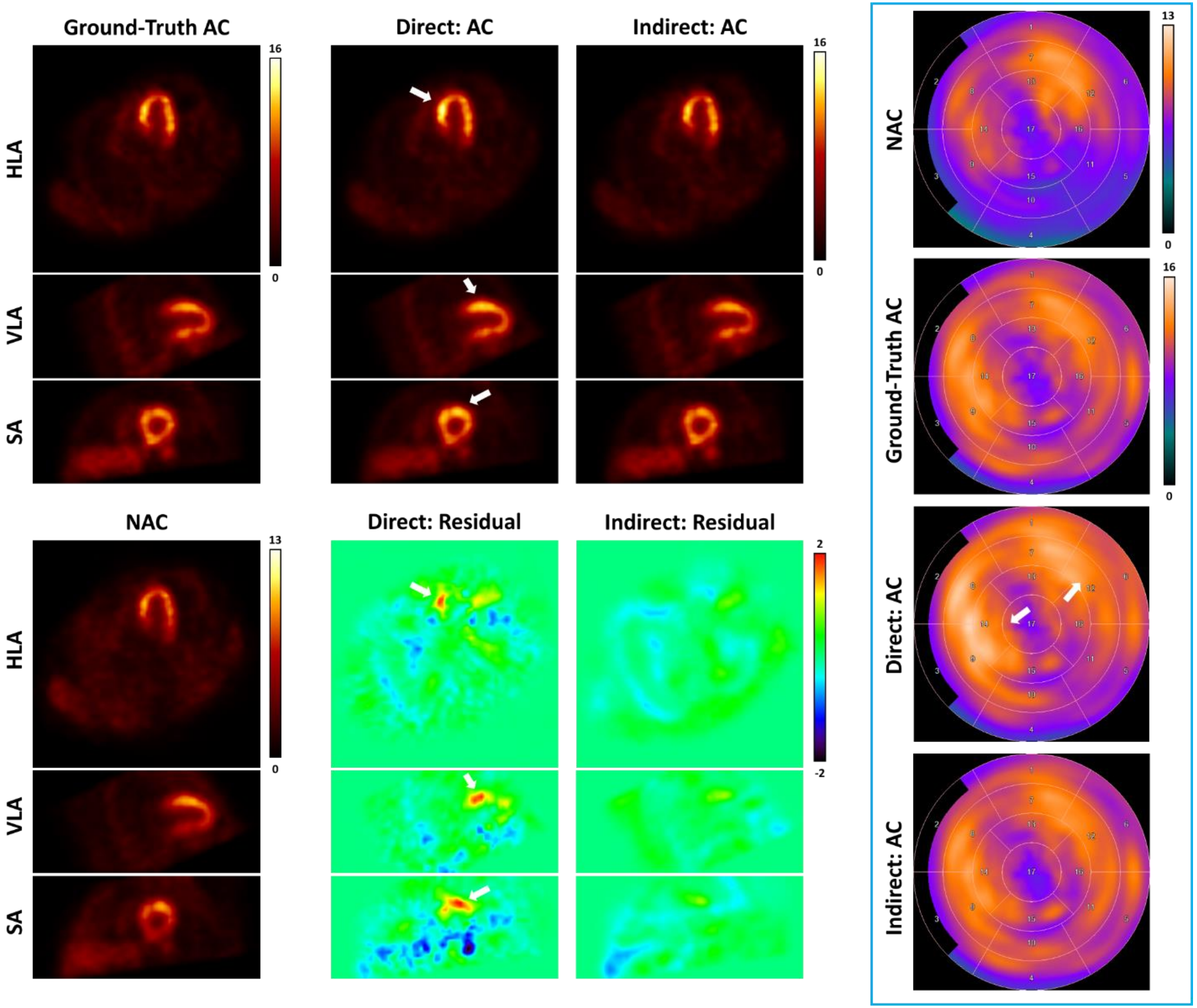
Over Correction



2 in 100 cases

Yang, Jaewon, et al. "Direct attenuation correction using deep learning for cardiac SPECT: A feasibility study." Journal of Nuclear Medicine (2021).

Comparison of Indirect and Direct Approaches for General Purpose SPECT

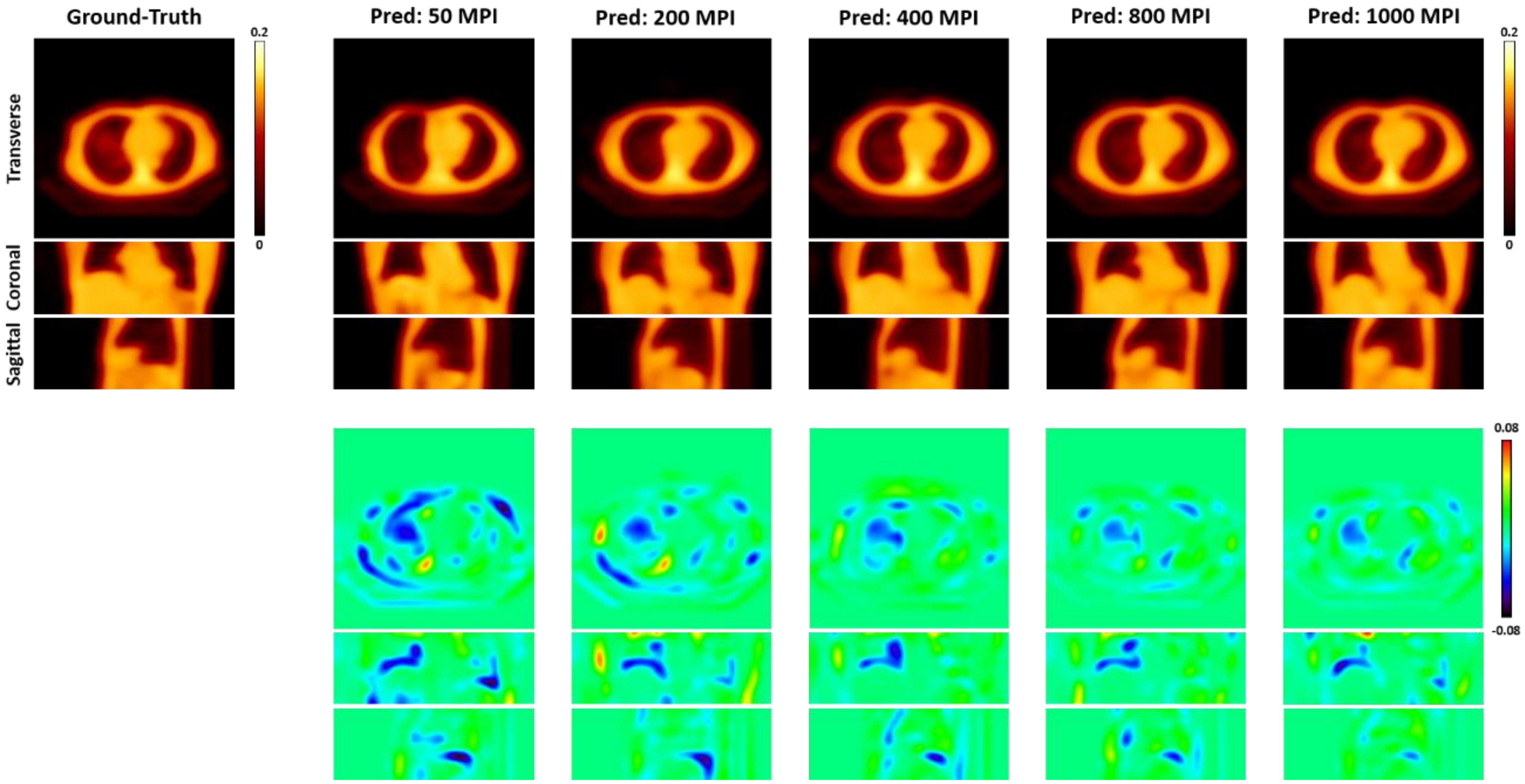


Chen, X.,*et al.* Direct and indirect strategies of deep-learning-based attenuation correction for general purpose and dedicated cardiac SPECT. *Eur J Nucl Med Mol Imaging* (Feb, 2022).

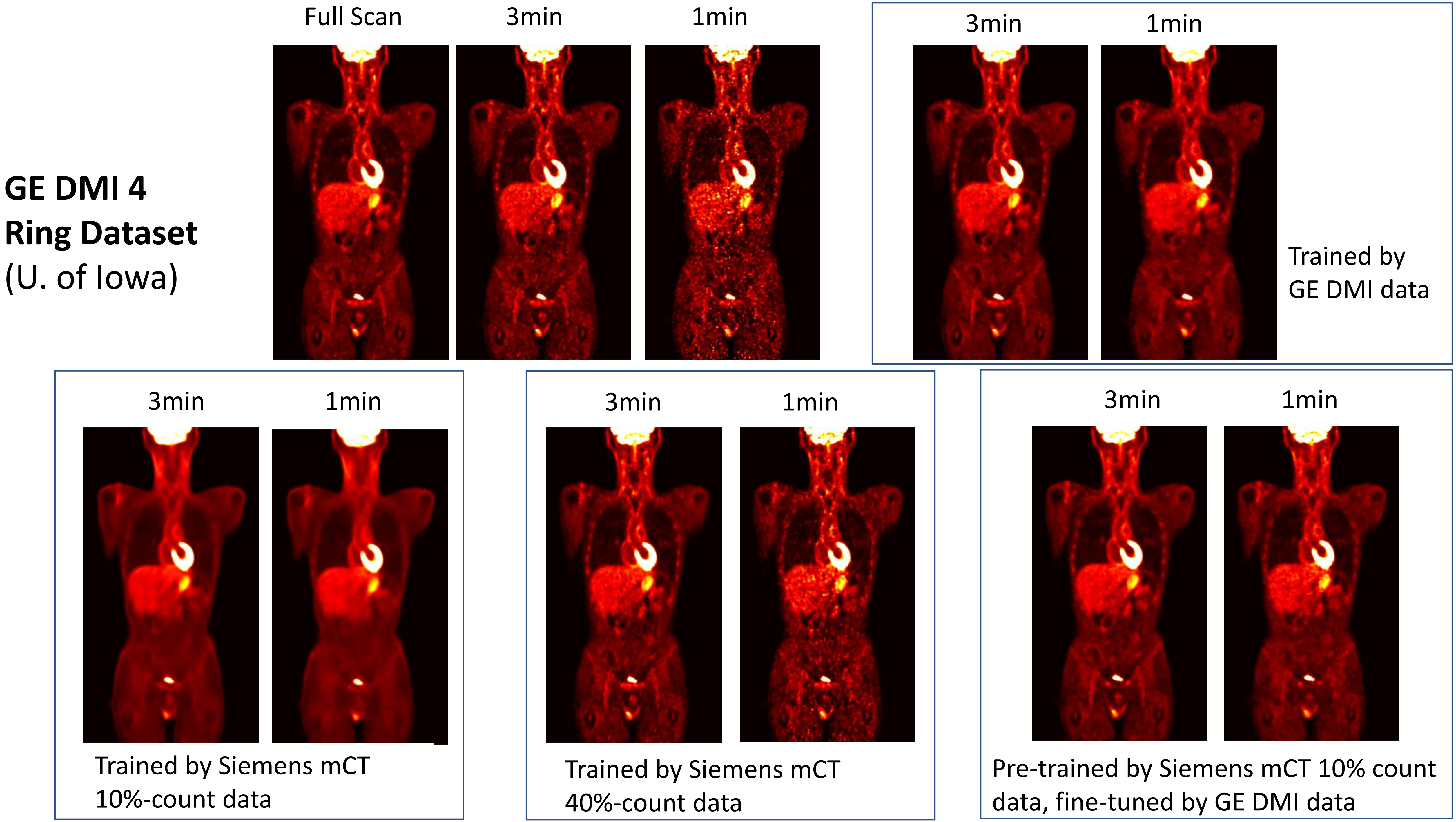
Questions to Panel 4: (FDA, CMS, and NIH)

- How much training datasets are needed?
- How diverse the training datasets need to be?
  - E.g. scanners, tracers, vendors
- How comprehensive the validations need to be?
  - E.g. patient population, disease types

How many data are needed for training?



Chen, X.,*et al.* Direct and indirect strategies of deep-learning-based attenuation correction for general purpose and dedicated cardiac SPECT. *Eur J Nucl Med Mol Imaging* (2022).

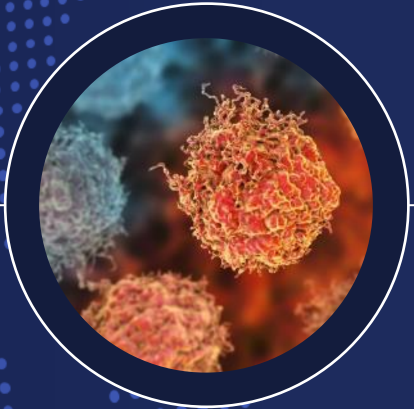

GE DMI 4
Ring Dataset
(U. of Iowa)
Full Scan
3min
1min
3min
1min
Trained by
GE DMI data
3min
1min
Trained by Siemens mCT
10%-count data
3min
1min
Trained by Siemens mCT
40%-count data
3min
1min
Pre-trained by Siemens mCT 10% count
data, fine-tuned by GE DMI data

Thoughts

- More upfront information related to translation can help data scientists develop more translatable AI technologies
- Have such information and considerations in the early phase of technology development

Lantheus

FIND > FIGHT > FOLLOW™



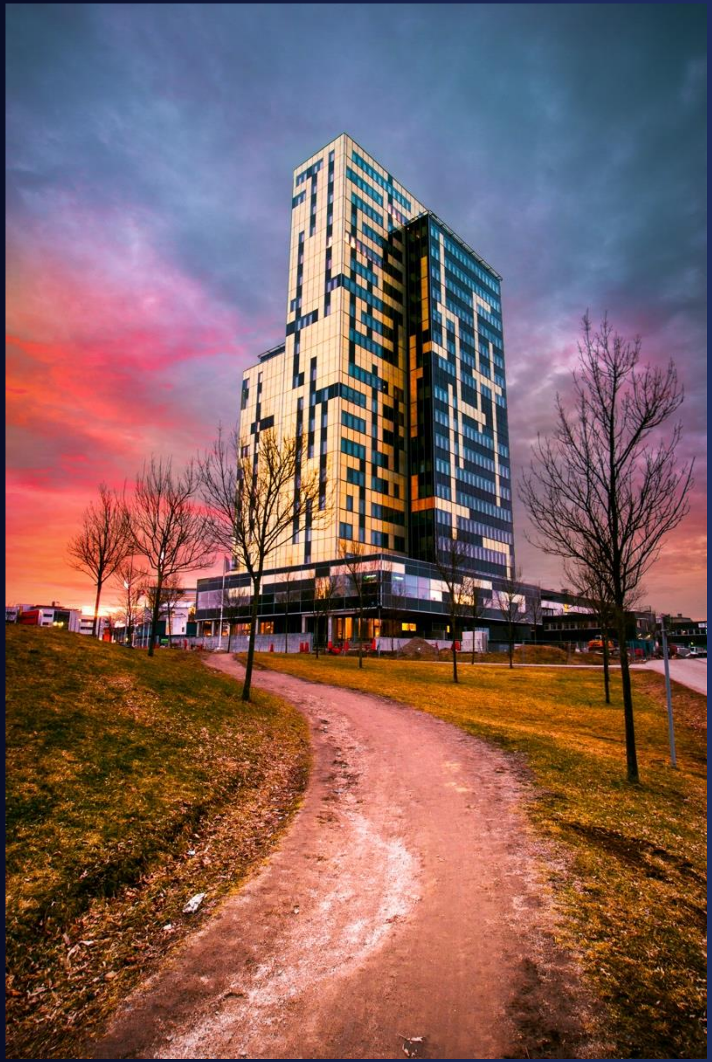
Too Much, Too Soon?

Reflections on EXINI's Path from Past to Present

Karl Sjöstrand

EXINI Diagnostics

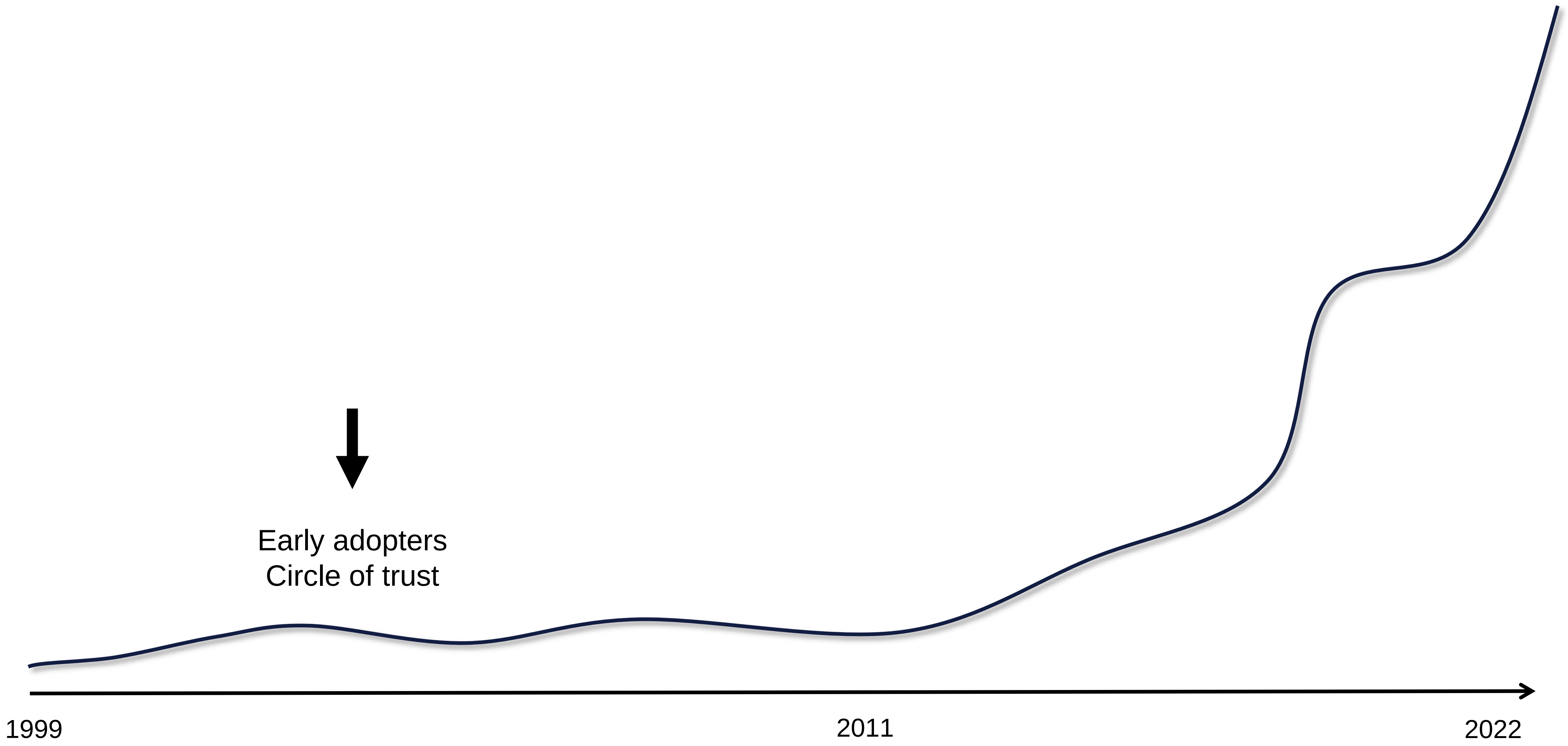
EXINI



About EXINI

- Business: Objective and standardized assessments from medical images for accurate staging, prognosis and treatment selection
- Incorporated in 1999: 23 years of experience in imaging biomarker software for nuclear medicine
- Based in Lund, Sweden
- EXINI Diagnostics AB is a wholly owned subsidiary of Lantheus Holdings.

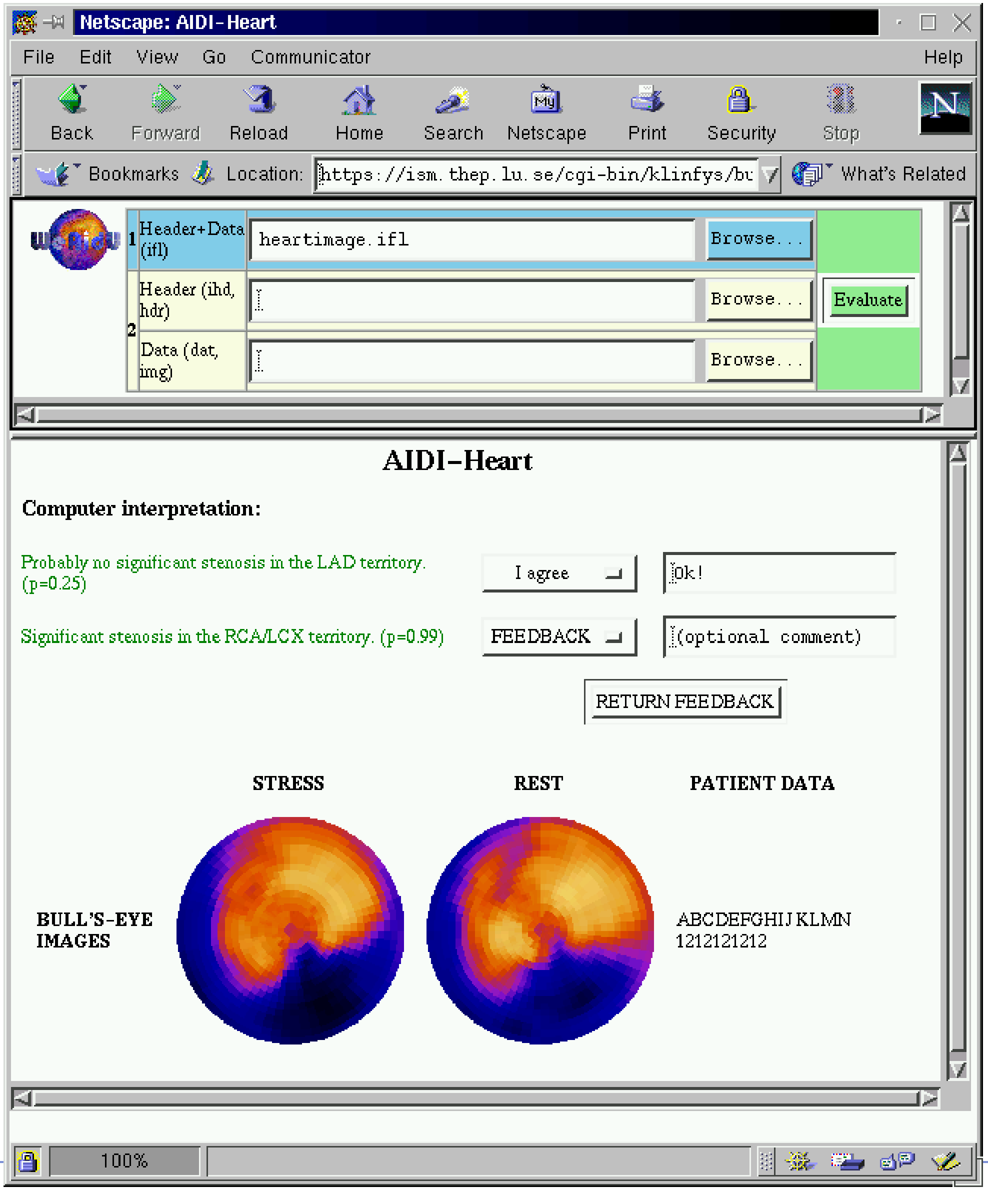
Timeline of Company Momentum



Timing is Everything

- Early 2000s
  - Cloud & browser based
  - Black box AI

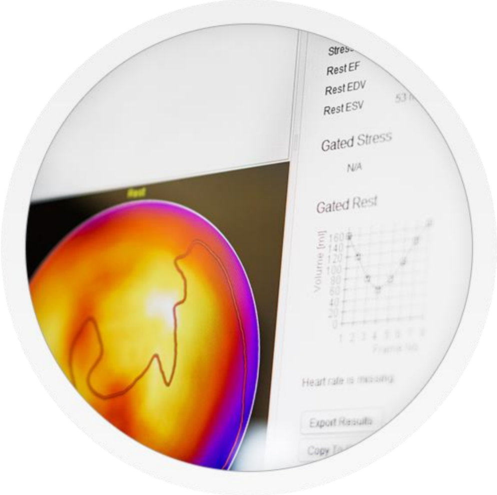
1. De-identify images
2. Send to cloud servers
3. Cloud AI processing
4. Receive results (diagnosis)
5. Re-identify & report



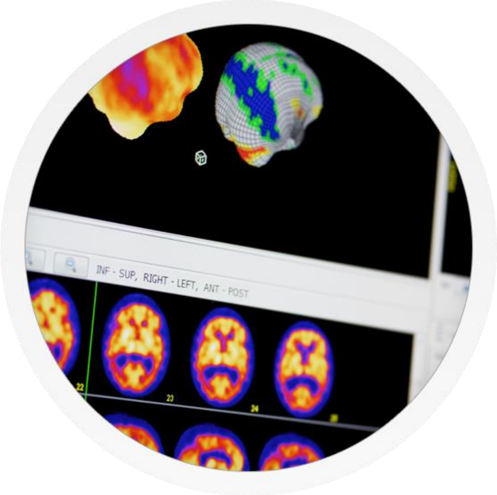
Course Correction

- Stand-alone applications
- Strong focus on clinical questions and unmet needs
- “Proximity effect” limited sales, inefficient sales process, small markets

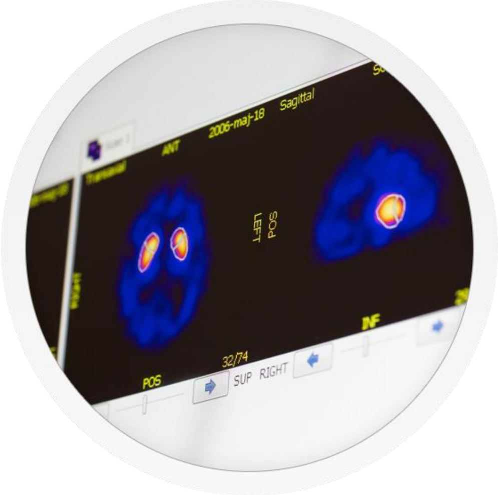
EXINI bone BSI



EXINI heart

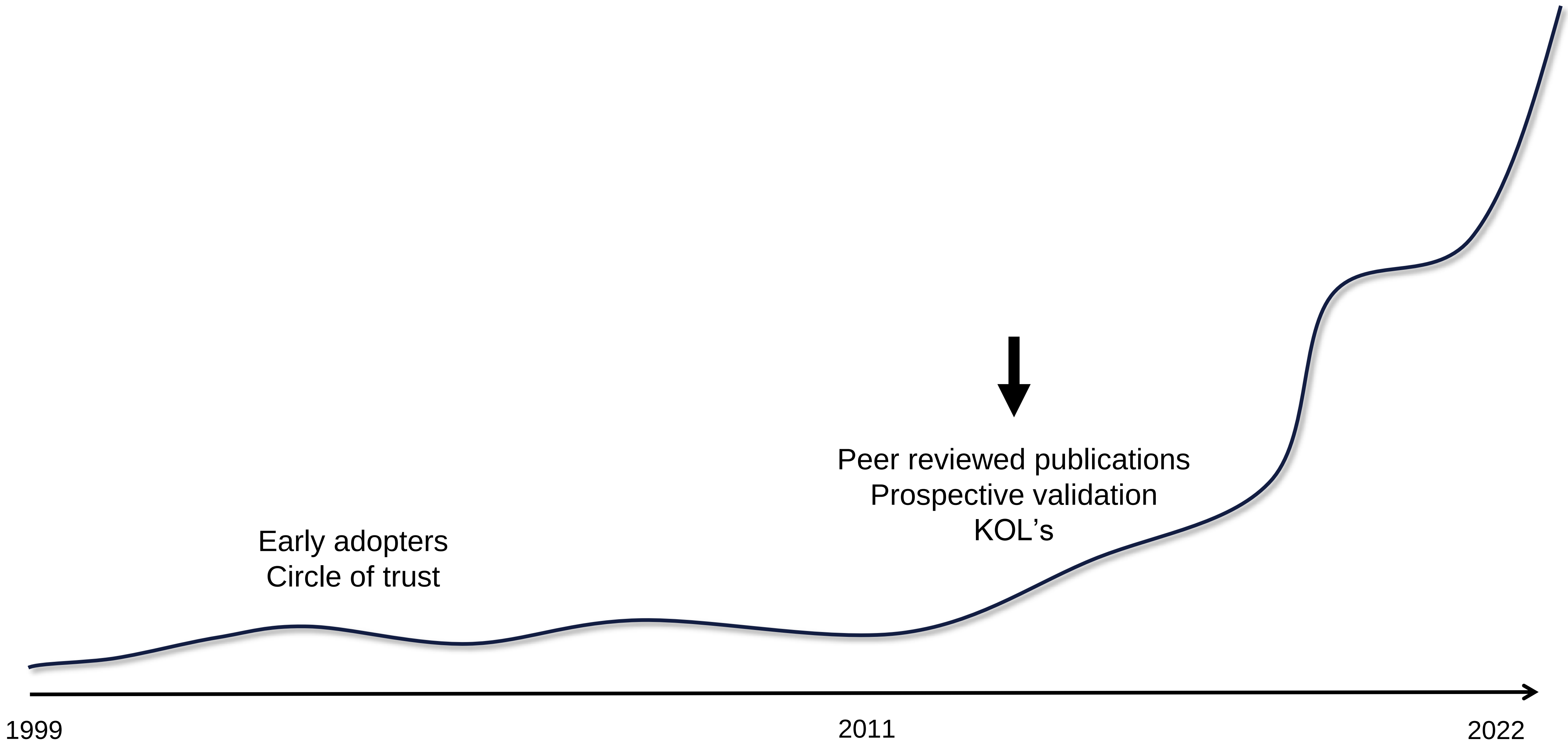


EXINI brain



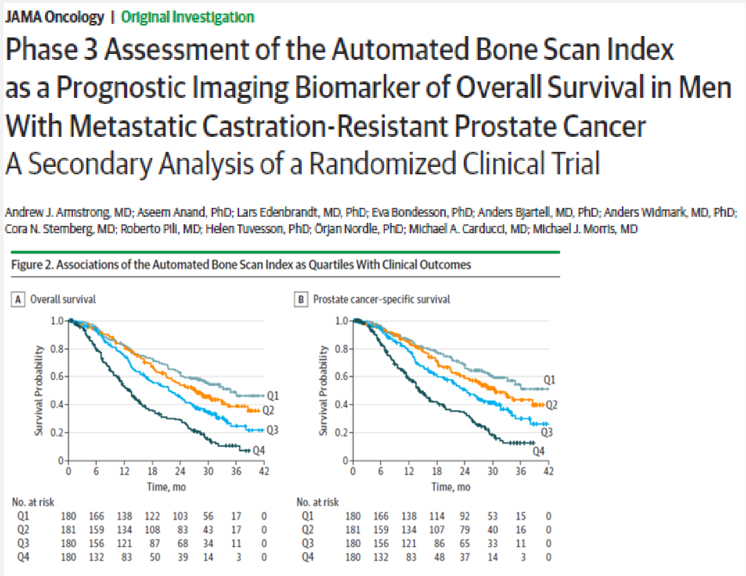
EXINI dat

Timeline of Company Momentum



aBSI

Automated Bone Scan Index (aBSI) as an independent **prognostic** biomarker for overall survival. (Armstrong, A. et al. JAMA Oncology 2018)


JAMA Oncology | Original Investigation

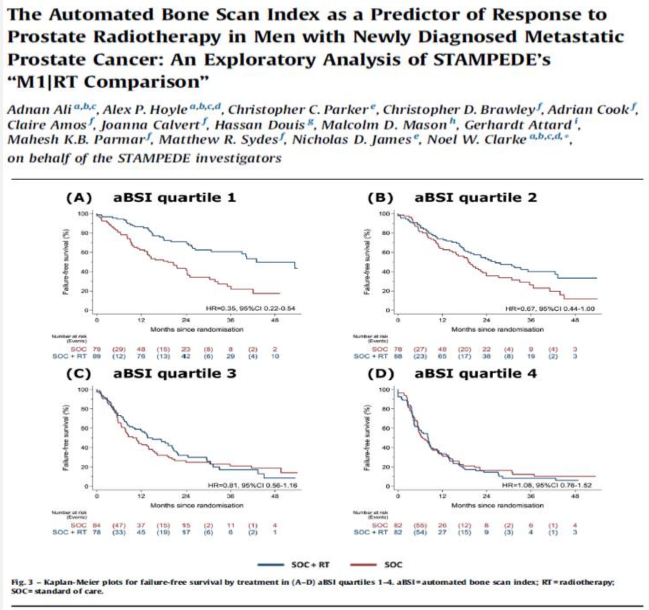
Phase 3 Assessment of the Automated Bone Scan Index as a Prognostic Imaging Biomarker of Overall Survival in Men With Metastatic Castration-Resistant Prostate Cancer

A Secondary Analysis of a Randomized Clinical Trial

Andrew J. Armstrong, MD; Aseem Anand, PhD; Lars Edenbrandt, MD, PhD; Eva Bondesson, PhD; Anders Bjartell, MD, PhD; Anders Widmark, MD, PhD; Cora N. Sternberg, MD; Roberto Pili, MD; Helen Tuvesson, PhD; Örjan Nordle, PhD; Michael A. Carducci, MD; Michael J. Morris, MD

Figure 2. Associations of the Automated Bone Scan Index as Quartiles With Clinical Outcomes

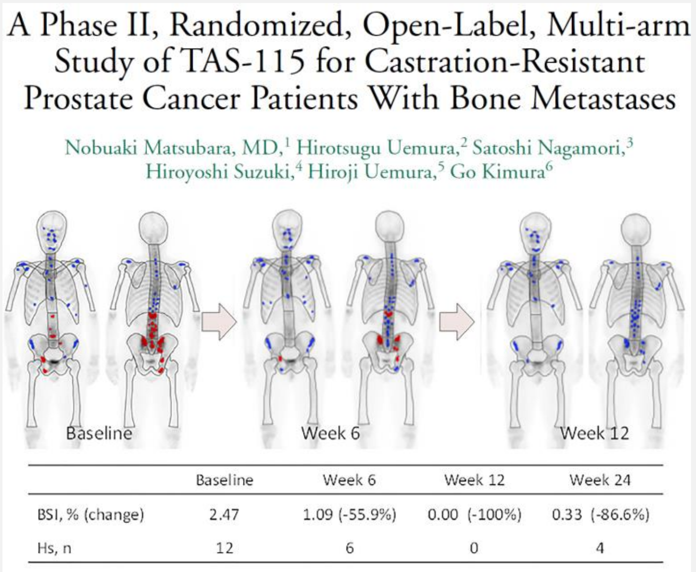
aBSI as a **predictor of response to prostate radiotherapy** in men with newly diagnosed metastatic prostate cancer. (Ali, A. et al. European Urol. 2019)


The Automated Bone Scan Index as a Predictor of Response to Prostate Radiotherapy in Men with Newly Diagnosed Metastatic Prostate Cancer: An Exploratory Analysis of STAMPEDE's "M1|RT Comparison"

Adnan Ali, Alex P. Hoyle, Christopher C. Parker, Christopher D. Brawley, Adrian Cook, Claire Amos, Joanna Calvert, Hassan Douis, Malcolm D. Mason, Gerhardt Attard, Mahesh K.B. Parmar, Matthew R. Sydes, Nicholas D. James, Noel W. Clarke, on behalf of the STAMPEDE investigators

Fig. 3 – Kaplan-Meier plots for failure-free survival by treatment in (A–D) aBSI quartiles 1–4. aBSI = automated bone scan index; RT = radiotherapy; SOC = standard of care.

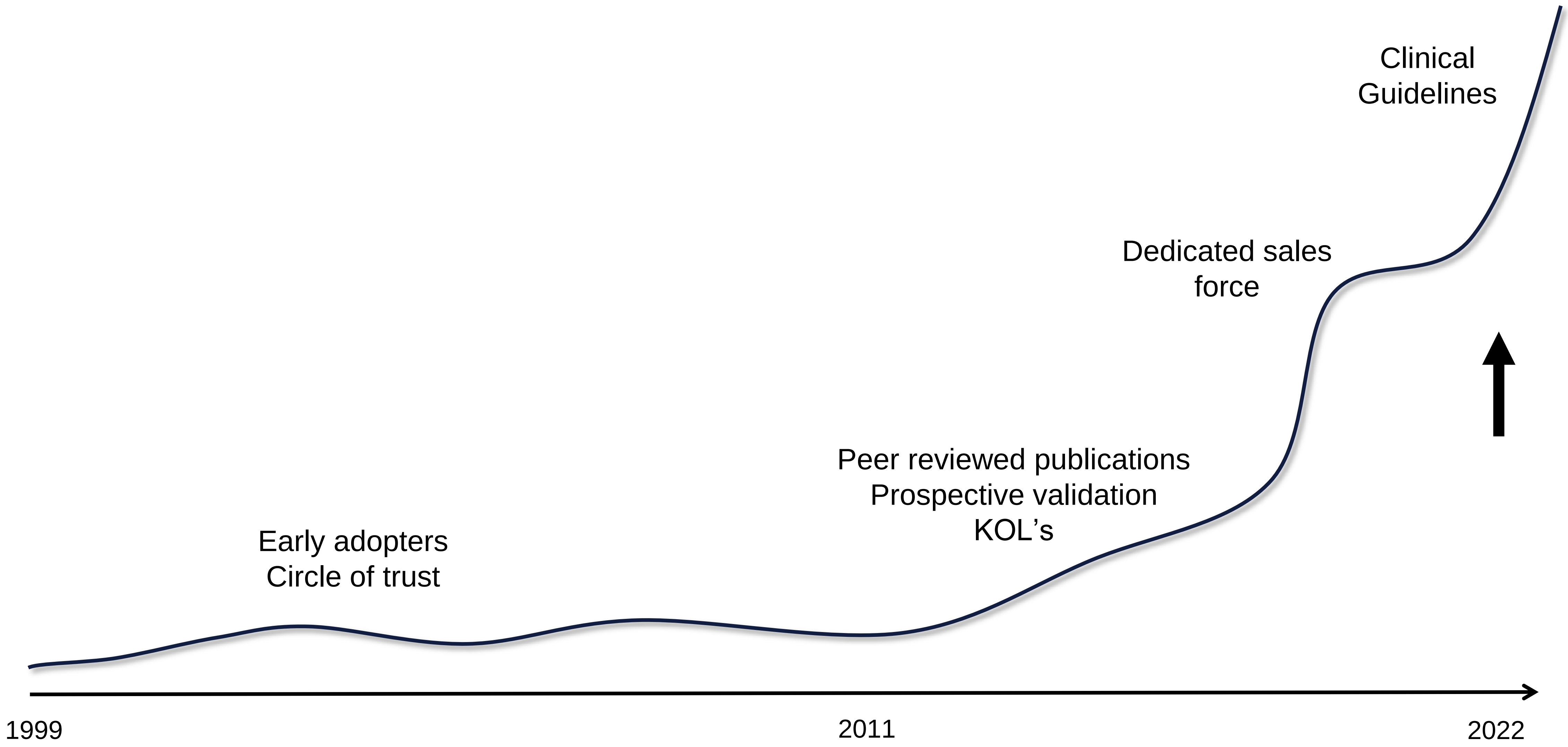
aBSI as **primary endpoint to evaluate efficacy** of TAS-115 as treatment for castration resistant prostate cancer with bone metastasis. (Matsubara, N. et al. Clinc. GU Cancer)


A Phase II, Randomized, Open-Label, Multi-arm Study of TAS-115 for Castration-Resistant Prostate Cancer Patients With Bone Metastases

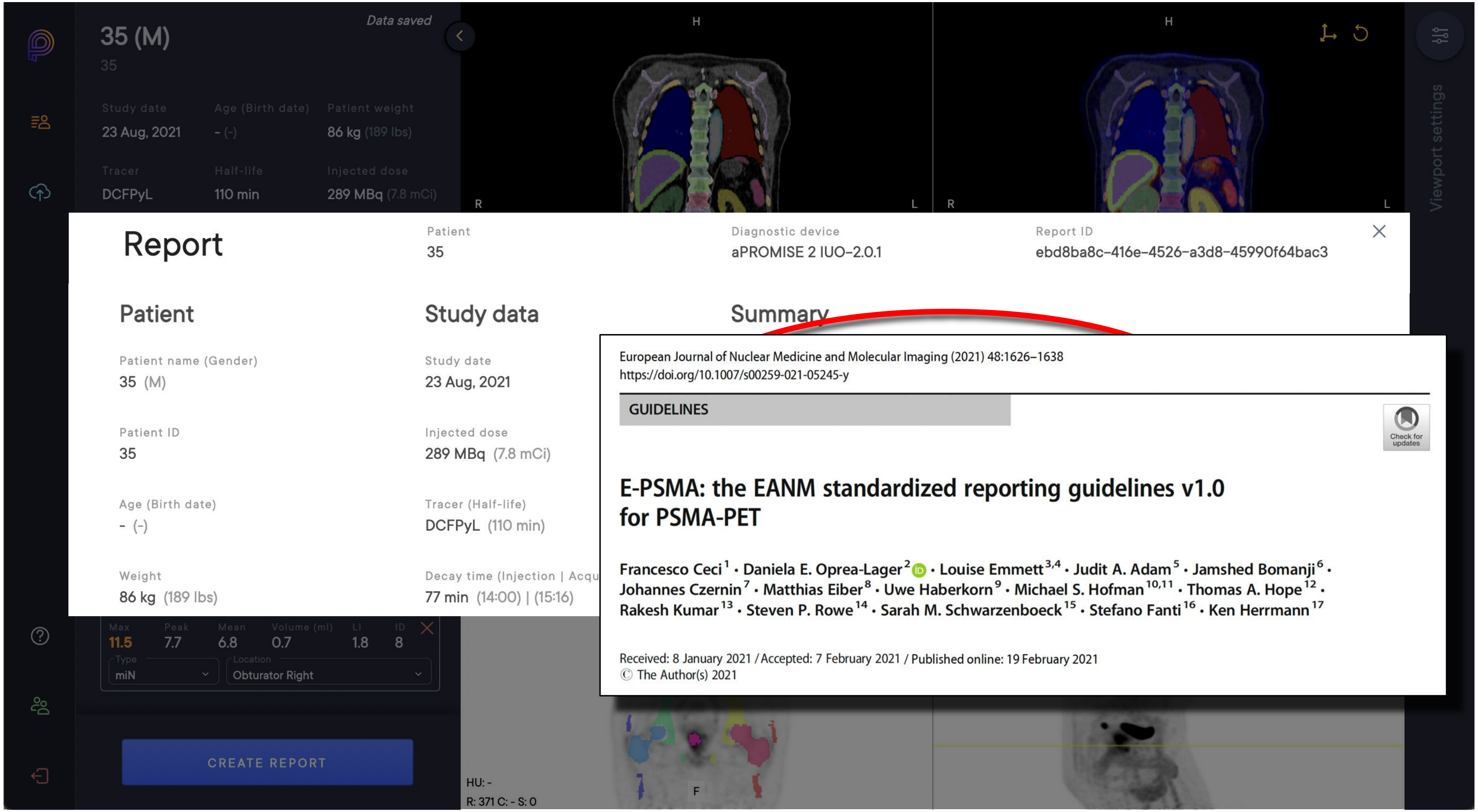
Nobuaki Matsubara, MD,[1] Hirotsugu Uemura,[2] Satoshi Nagamori,[3] Hiroyoshi Suzuki,[4] Hiroji Uemura,[5] Go Kimura[6]

	Baseline	Week 6	Week 12	Week 24
BSI, % (change)	2.47	1.09 (-55.9%)	0.00 (-100%)	0.33 (-86.6%)
Hs, n	12	6	0	4

Timeline of Company Momentum



PYLARIFY AI

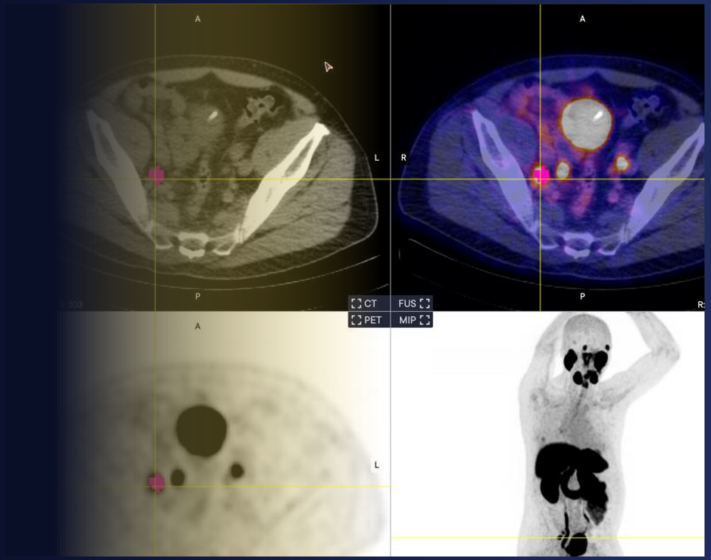


35 (M)
35
Data saved
Study date
23 Aug, 2021
Age (Birth date)
- (-)
Patient weight
86 kg (189 lbs)
Tracer
DCFPyL
Half-life
110 min
Injected dose
289 MBq (7.8 mCi)
H
R
L
Viewport settings
Report
Patient
35
Diagnostic device
aPROMISE 2 IUO-2.0.1
Report ID
ebd8ba8c-416e-4526-a3d8-45990f64bac3
Patient
Patient name (Gender)
35 (M)
Patient ID
35
Age (Birth date)
- (-)
Weight
86 kg (189 lbs)
Study data
Study date
23 Aug, 2021
Injected dose
289 MBq (7.8 mCi)
Tracer (Half-life)
DCFPyL (110 min)
Decay time (Injection | Acqu
77 min (14:00) | (15:16)
Summary
Max 11.5
Peak 7.7
Mean 6.8
Volume (ml) 0.7
LI 1.8
ID 8
Type
miN
Location
Obturator Right
CREATE REPORT
HU: -
R: 371 C: - S: 0
F
European Journal of Nuclear Medicine and Molecular Imaging (2021) 48:1626–1638
https://doi.org/10.1007/s00259-021-05245-y
GUIDELINES
Check for updates
E-PSMA: the EANM standardized reporting guidelines v1.0 for PSMA-PET
Francesco Ceci 1 · Daniela E. Oprea-Lager 2 · Louise Emmett 3,4 · Judit A. Adam 5 · Jamshed Bomanji 6 · Johannes Czernin 7 · Matthias Eiber 8 · Uwe Haberkorn 9 · Michael S. Hofman 10,11 · Thomas A. Hope 12 · Rakesh Kumar 13 · Steven P. Rowe 14 · Sarah M. Schwarzenboeck 15 · Stefano Fanti 16 · Ken Herrmann 17
Received: 8 January 2021 / Accepted: 7 February 2021 / Published online: 19 February 2021
© The Author(s) 2021

Status

Currently rolling out to PYLARIFY PSMA customers in US and EU

- Enabling reproducible, standardized and quantitative reporting in PSMA PET/CT



Challenges

- Security audit is not standardized
- Deep integration with existing clinical workflow is essential
  - But it is different in every site
  - No user interface standard
- Open standards such as DICOM and HL7 are great - but conformance is highly variable
- Cloud deployments are straight forward and safe – but largely not accepted
- Local installations are time consuming to set up, difficult to set up, monitor and maintain

AI in Nuclear Medicine

Prepare Like You're Training for a Triathlon

Tim Adams, MIM Software
Nuclear Medicine Market Director

A Challenge with 3 Phases

PHASE 1

Design

PHASE 2

Analyze

PHASE 3

Implement

Phase 1: **DESIGN**

Challenges

CLINICIANS

- Time for data creation
- Secure tools to share data
- Report anonymization

INDUSTRY

- Data access
- Clean, multi-institutional datasets (unbiased)
- Data agreements

Phase 1: **DESIGN**

Needs

SHARED

- “Neutral Ground” data repositories
- Increased access to trial data
- Community registries
- Data sharing education

Phase 2: **ANALYZE**

Challenges

CLINICIANS

- Unclear acceptance criteria
- Testing AI is disruptive to workflow
- Experience gaps

INDUSTRY

- Unclear regulatory landscape
- Validation requires clinical support
- Performance metrics are unclear
- Challenging to evaluate the Human-AI Team (Guideline 7)*

*https://www.fda.gov/media/153486/download

Phase 2: **ANALYZE**

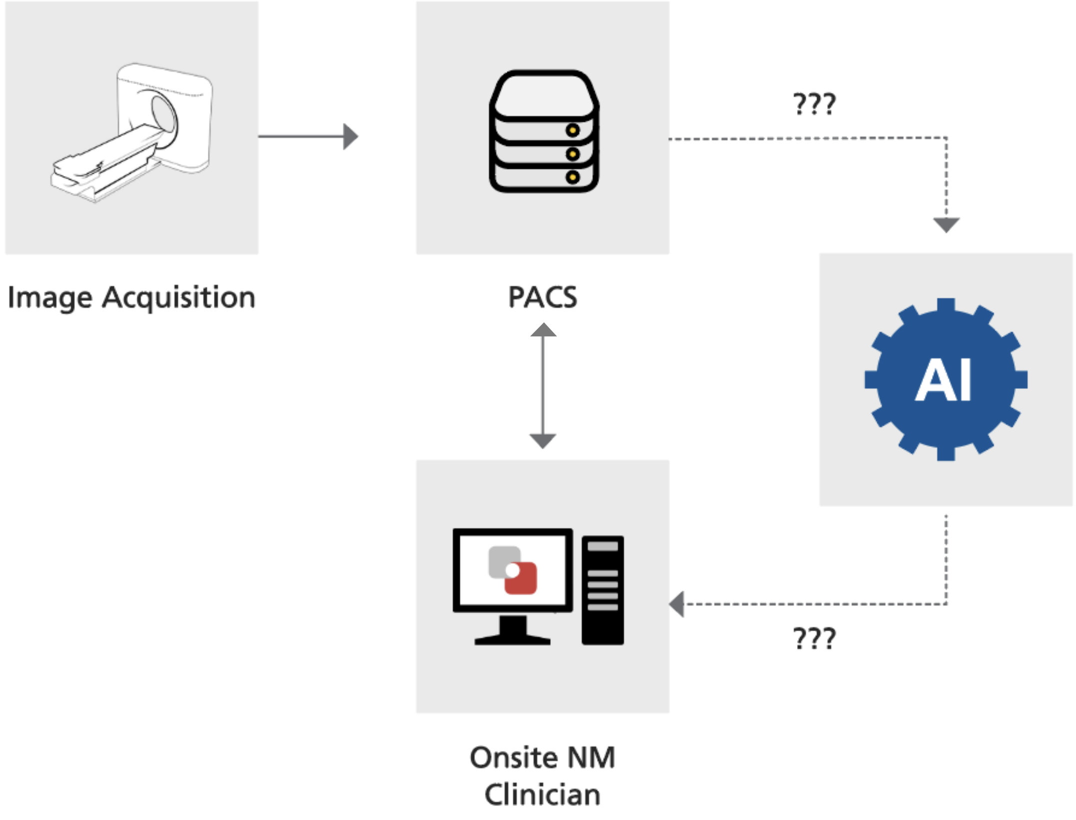
Needs

SHARED

- FDA/clinician performance standards
- Finalized regulatory guidelines
- Expert ground truth data
- Integrated AI testing ecosystems
- Workgroups to support AI validation

510(k) Cleared

How does a clinician validate AI?



Phase 3: **IMPLEMENT**

Challenges

CLINICIANS | **INDUSTRY**

- Lack of education resources about AI
- IT support and limitations of current hardware
- Small vs. large hospitals have different deployment issues
- Integrations into the clinical workflow
- Incentives and reimbursement
- Performance monitoring

"...a lack of education and training about AI could limit the technology from achieving its full potential."

- **Samantha Santomartino, Dr. Paul Yi**
  University of Maryland

Yee Madden, Kate. "**Both radiologists and medical students see the value of AI**" *AuntMinnie.com*, https://bit.ly/3wfILS8. 2 February 2022.

Phase 3: **IMPLEMENT**

Needs

SHARED

- Education on AI as part of residency curriculum
- Non-disruptive quality checks
- Prioritization of IT resources
- Society support for cloud technologies
- Pathways and guidance to establish where AI could fit in fee structures
- Guidance from clinicians/FDA on performance monitoring

$ SET FEE STRUCTURE

Summary

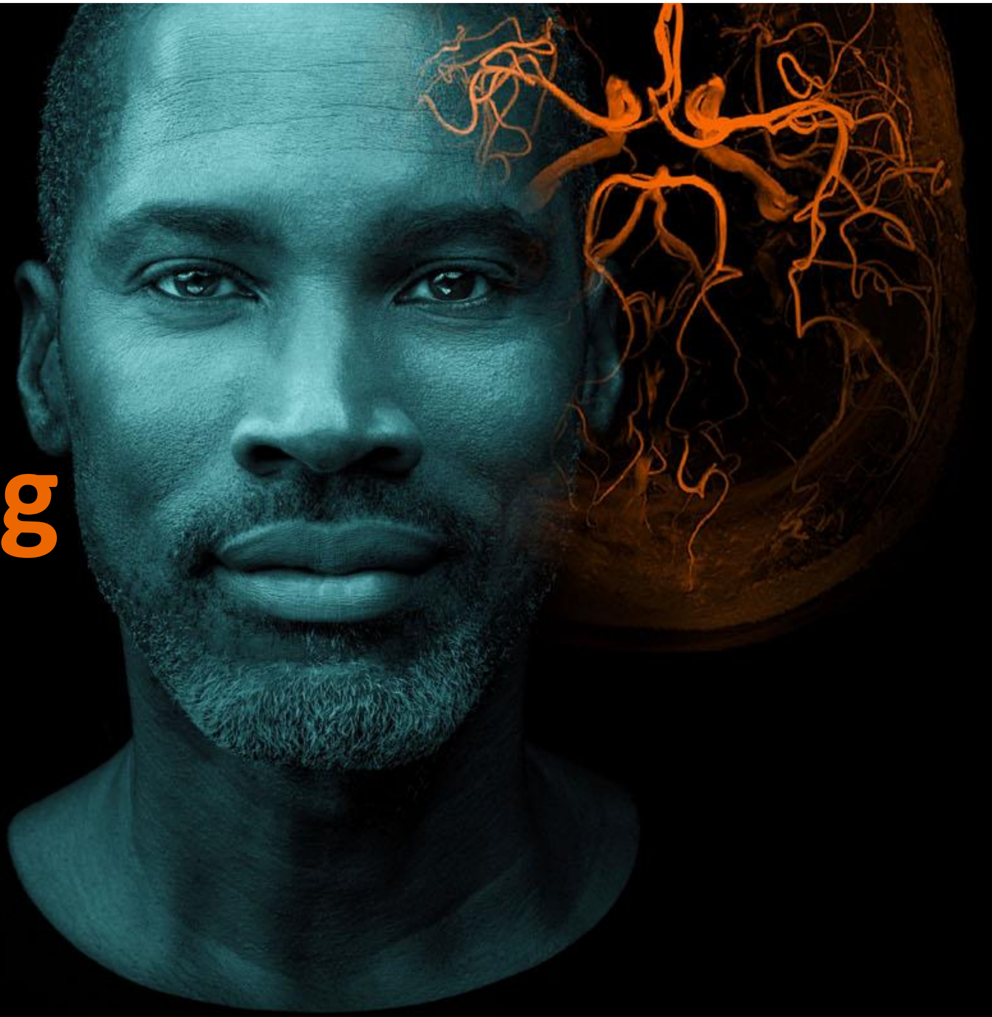
INDUSTRY NEEDS:

- Data access is critical, but education and resources on how to support data initiatives are lacking. We need resources like the Data Science Institute website to inform, connect, and educate the community on AI
- Acceptance criterias and performance standards to streamline approvals
- FDA guidance around topics, including performance monitoring
- Strengthening relationships between clinicians and hospital IT
- Workgroups and consortiums around specific clinical problems
- Clarity regarding reimbursement for AI

MIM
SOFTWARE™

SIEMENS Healthineers

AI in Molecular Imaging



Sven Zuehlsdorff, Ph.D.
Sr. Director, Research
Siemens Healthineers, Molecular Imaging

21 March 2022

AI in Molecular Imaging: Selected use cases

SIEMENS Healthineers

1. Innovate Modality Business

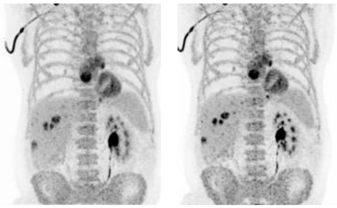
Image Formation

- Data driven gating
- Kinetic Modeling
- PACS ready image preparation
- Low count image reconstruction
- Denoising
- PET/CT image registration

Patient Workflow

- Patient positioning
- Scan planning
- Device less Gating
- Low dose/fast scan
- Breath hold scan

Example



Data driven gating to reduce impact of motion.

2. Expand Diagnostic Offerings

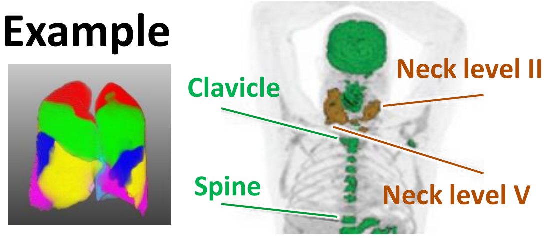
Efficient Quantification

- Segmentation: lesions, organs
- PERCIST, PROMISE, Deauville, …
- Neuro data base comparison
- Radiomics

Diagnostic Tools

- Lung V/Q on lobe/segments level
- Lesion classification
- Disease staging support
- Differential diagnosis in neuro

Example



AI to parcellate lung lobes/segments.
AI classifies normal vs. suspicious.

3. Lead Clinical Decisions

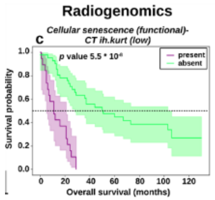
Therapy selection, monitoring

- Theranostics, dosimetry, monitoring
- Emerging therapies
- Auto Staging: Onco/neuro
- Risk stratification, phenotyping

New Frontiers

- Early diagnosis / screening
- Virtual biopsy
- Predictive disease modeling
- System biology/organ networks

Example



Combination of radiomics and genomics outperforms prognostic value of genetics and imaging markers alone.

AI in Molecular Imaging: Opportunities and challenges (1/2)

SIEMENS Healthineers

Access to Data
- Access to data
- Quality
- Diversity
- Patient privacy
- Cost of data

Foundation

Algorithm
- Technology
- Model/Learning
- Training
- Testing
- Validation

Integrity

Evaluation
- Concept
- Technical
- Clinical
- Real world
- Post market

Quality

Product
- R&D invest
- Deployment
- Market
- Maintenance
- Viability

Access

Regulatory
- Global market
- Local clearance
- Claims/Evidence
- Clinical Trials
- GMLP

Safety

AI in Molecular Imaging: Opportunities and challenges (2/2)

Artificial Intelligence in **Molecular Imaging**
may be used to assist in
deriving clinically relevant and actionable information
in a fashion that is
Safe: does not cause harm, high in quality
Quantitative: **accurate, reproducible, robust**
Efficient: **automated, operator independent**

*"About **100 years** ago, **electricity transformed** every major **industry**. **AI** has advanced to the point where it **has the power to transform every major sector** in the coming years."*
Dr. Andrew Ng, Stanford University (2017)

*"There's **hype** about **artificial intelligence**, but most of the approaches suffer from **poor data quality** or not enough data. If you want to use **AI** as an **expert system**, to train people, to support people in their decisions, you have to make sure that the data**, the ground truth**, **is not wrong** from the beginning.*
Dr. Michael Schäfers, University of Münster

*"**AI**, if it's truly **meaningful**, needs to be **almost invisible**. **Don't change** the reader's **method**—**support it, add to it, augment it**, but don't change it."*
Dr. Carl von Gall, Siemens Healthineers

https://knowledge.wharton.upenn.edu/article/ai-new-electricity/
Fahmy S. *How we developed AutoID*, Imaging Life 2021, white paper, Siemens Healthineers, https://www.siemens-healthineers.com/molecular-imaging/news
The products/features (mentioned herein) are not commercially available in all countries and/or for all modalities. Future availability cannot be guaranteed.

Siemens Healthineers
Molecular Imaging

Siemens Medical Solutions USA, Inc.
810 Innovation Dr
Knoxville, TN 37932

Phone: +1 (865) 218-2000
siemens-healthineers.com

Sven Zuehlsdorff, Ph.D.
Sr. Director, Research

Siemens Medical Solutions USA, Inc.
2501 North Barrington Road
Hoffman Estates, IL 60192, USA

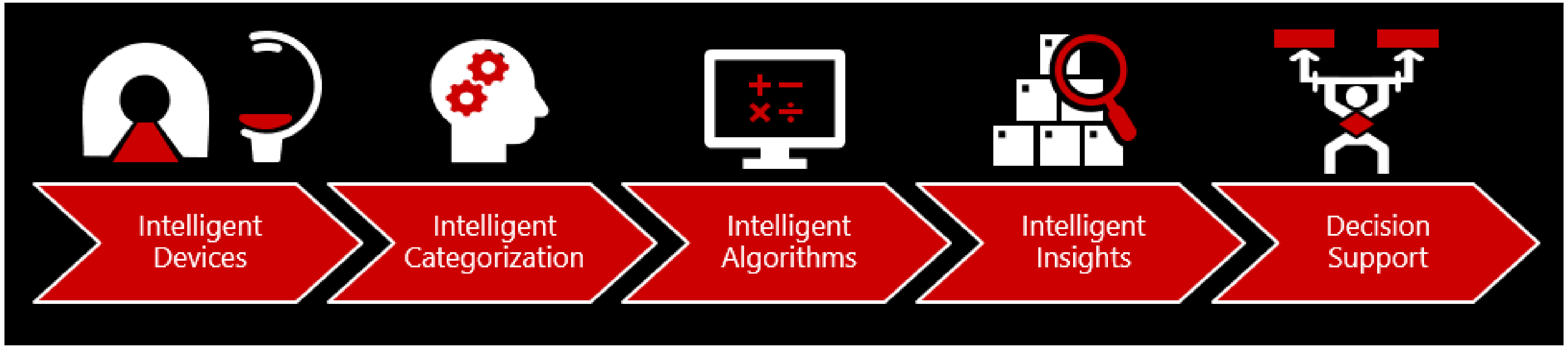
Mobile: +1 (773) 351-9496
sven.zuehlsdorff@siemens-healthineers.com

AI Applications at Canon Medical Research USA

Vernon Hills, IL

Evren ASMA
PET Image Reconstruction & Physics

CMRU Vision for Artificial Intelligence



- **Intelligent Devices**

  e.g. Scanners which know which region to scan for how long

- **Intelligent Categorization**

  e.g. Datasets/images with high probability of having lesions and lesion locations

- **Intelligent Algorithms**

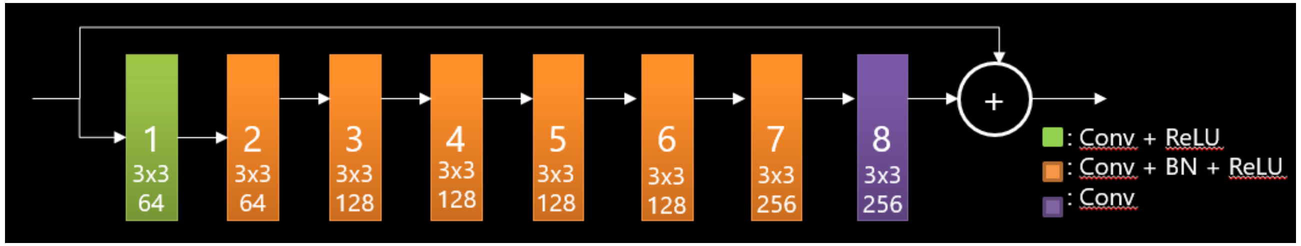
  e.g. AI-based denoising or AI-inside-the-recon or AI for corrections

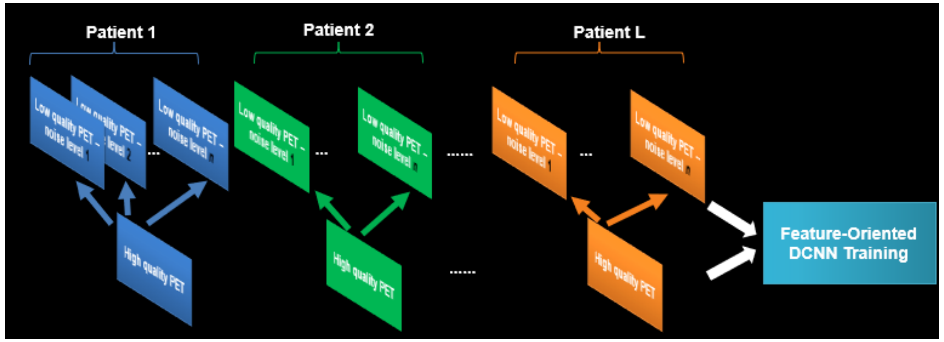
- **Decision Support**

  e.g. Assisting doctors in clinical decisions

AI Application: PET Image Denoising

- 8-layer residual deep convolutional network
- Approach can generate low noise images from input images ranging from low to very high noise
  - Trained with multiple noise levels
- Feature-oriented training weights features of interest higher during training
- Significantly improved quantitation over OSEM due to similar contrast levels but much lower noise





OSEM +GF

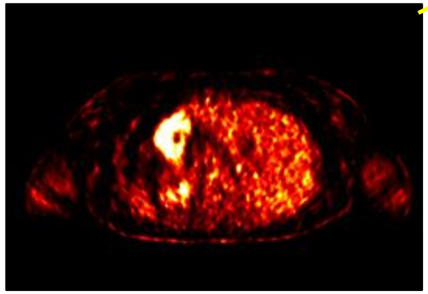
AiCE

Advanced Iterative Clear-IQ Engine

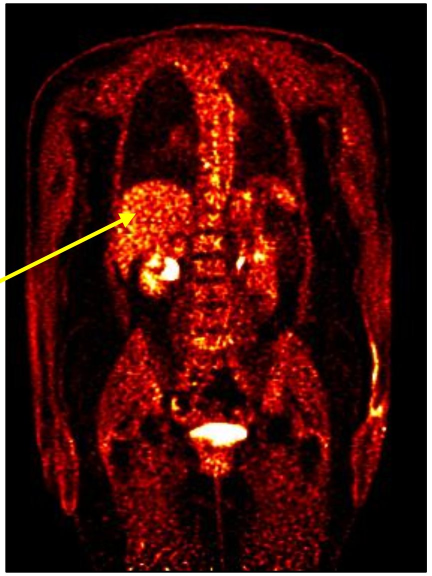
Chan, C., et al (2018). Noise Adaptive Deep Convolutional Neural Network for Whole-Body PET Denoising. *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2018 - Proceedings*. https://doi.org/10.1109/NSSMIC.2018.8824303

AI-Based Denoising Examples

Std = 30%



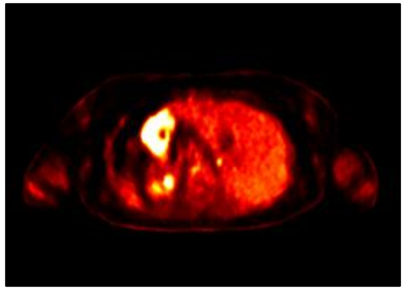
OSEM + GF



AiCE

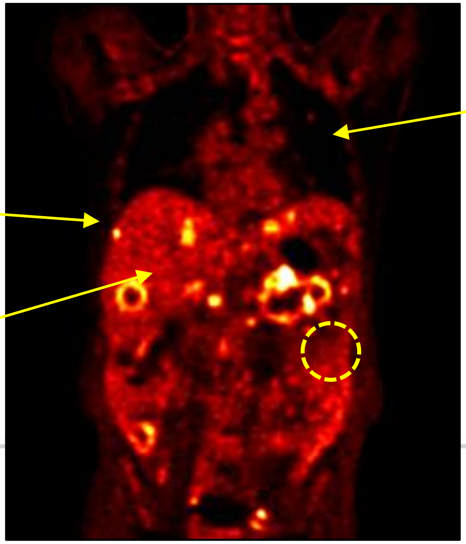
BMI: 29.7
PET Dose: 13.8 mCi
Glucose: 84 mg/dl
PET uptake time: 92
Scan time: 2.5 min /bed
Recon: OSEM 4 iter, 10ss, GF

Std = 6%



Significant noise reduction with AiCE compared to OSEM + Gaussian post-filtering

OSEM+GF:4 min/bed



CR = 29%

SUV_{mean} = 4.1

Std = 11%

AiCE:2 min/bed

CR = 48%

SUV_{mean} = 5.4

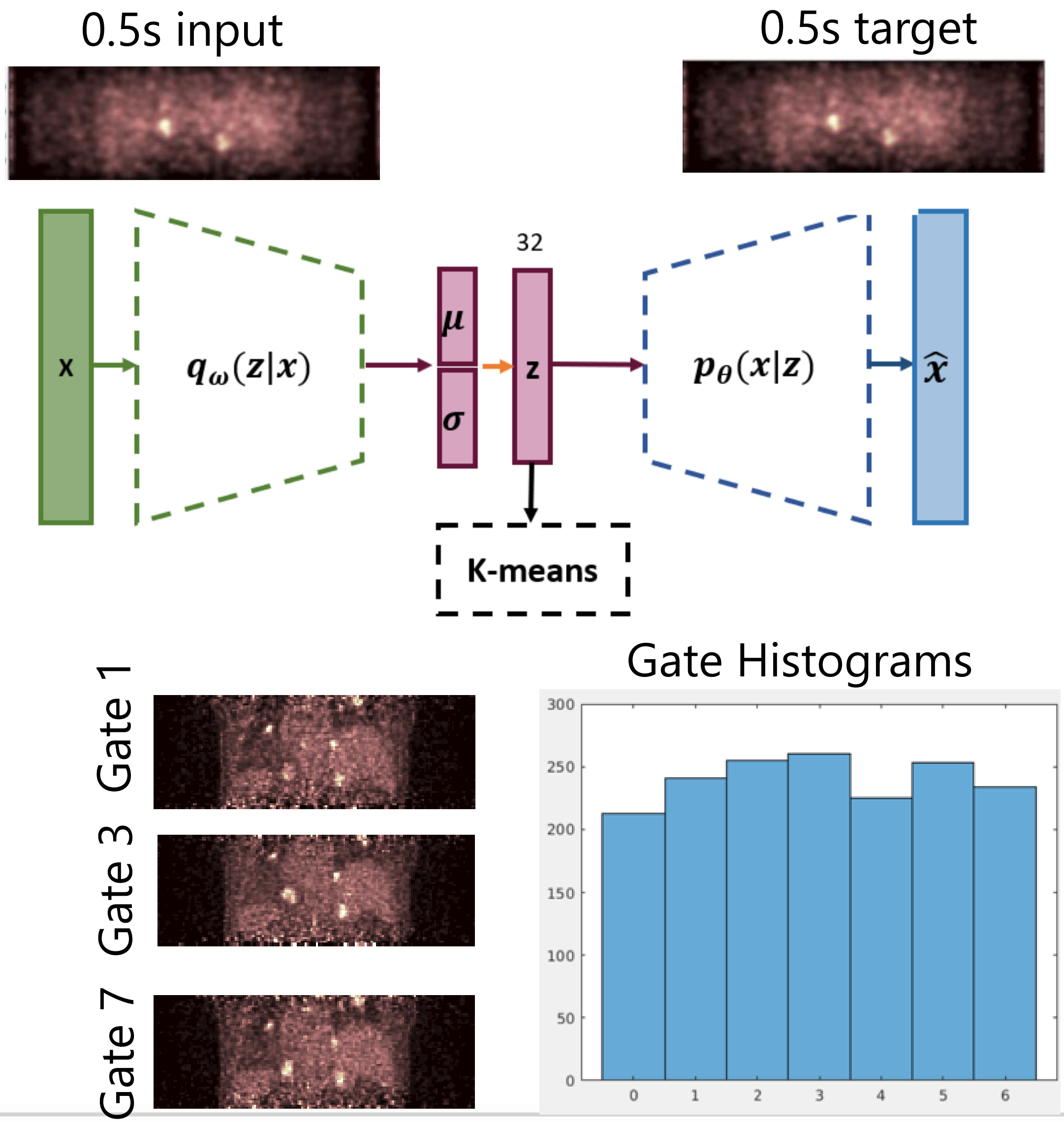
Std = 7%

Improved contrast-to-noise ratios in half the scan time

Canon

AI Application: Network-Based Data-Driven Gating

- Neural-network-based data driven gating clusters very short (0.5 sec) scan segments based on their network features
  - One could also apply PCA on network features to generate network-based gating signals for users
  - No optical or pressure-sensing external motion trackers are used
  - The result of clustering network features is "AI-gating" – not displacement or phase gating

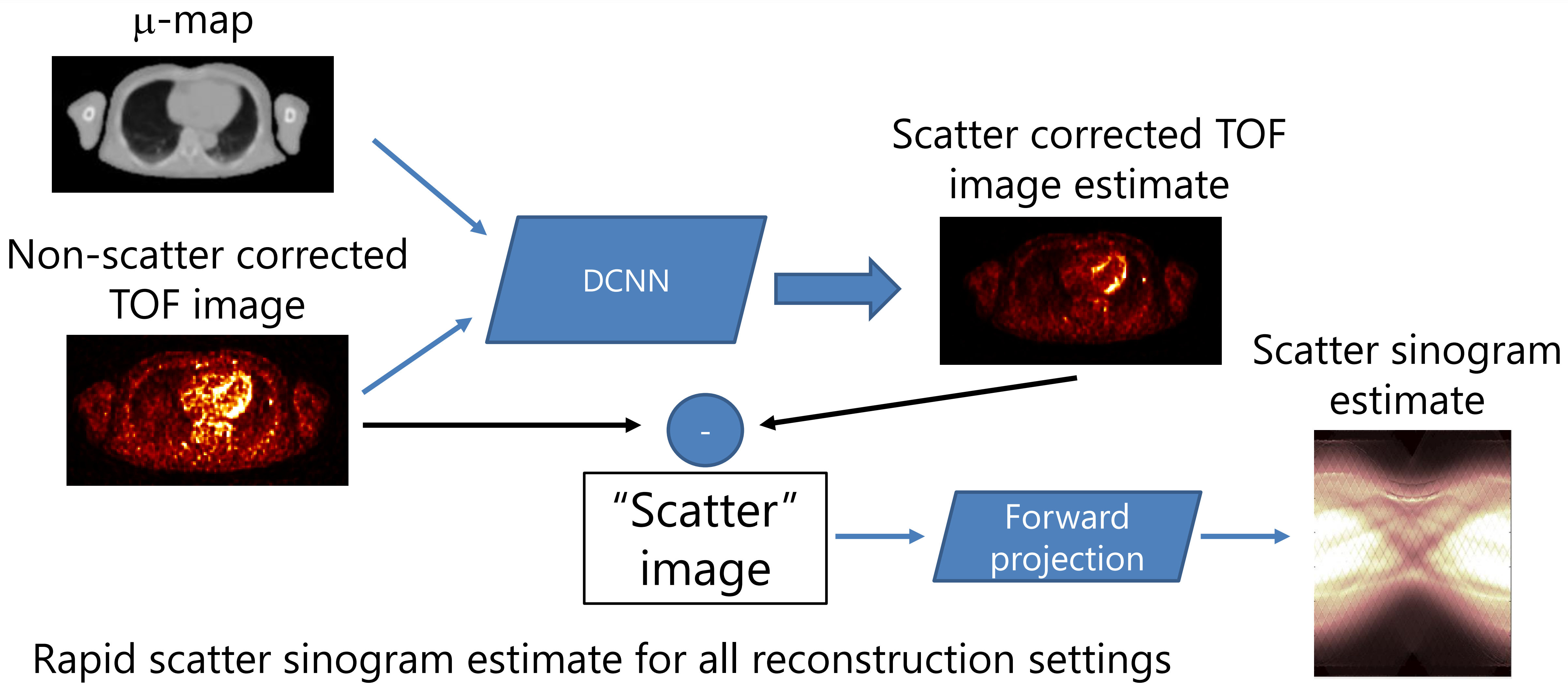


UC Davis – CMRU collaboration

Canon

Li, T., Zhang, M., Qi, W., Asma, E., & Qi, J. (2020). Motion correction of respiratory-gated PET images using deep learning based image registration framework. Physics in Medicine and Biology, 65(15). https://doi.org/10.1088/1361-6560/ab8688

AI Application: Network-Based Scatter Correction



- Rapid scatter sinogram estimate for all reconstruction settings
- Avoids complicated physical and mathematical modeling

UC Davis – CMRU collaboration

Canon

Xie, Z., Qi, W., Zhang. X.., Li T., Spencer B.A., Leung E.K., Wang G., Badawi R.D., Cherry S.R, Asma E., Qi J., (2022). Hybrid scatter correction for PET image reconstruction, SNMMI Annual Meeting submission

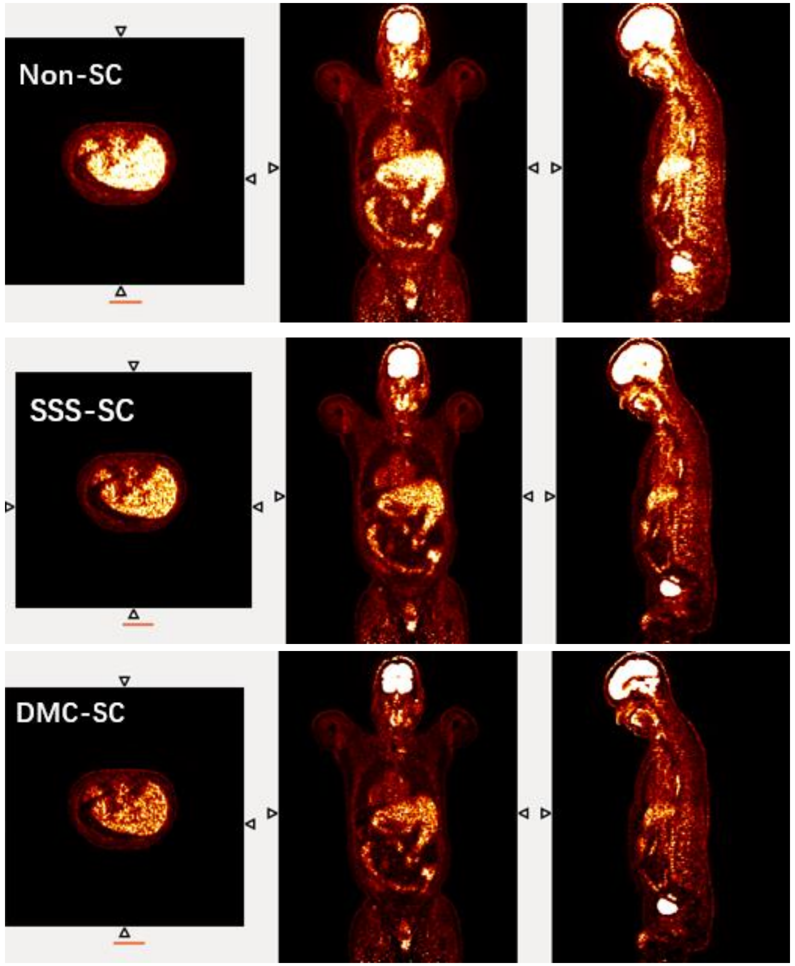
AI-Based Scatter Correction Examples

Non-scatter corrected image

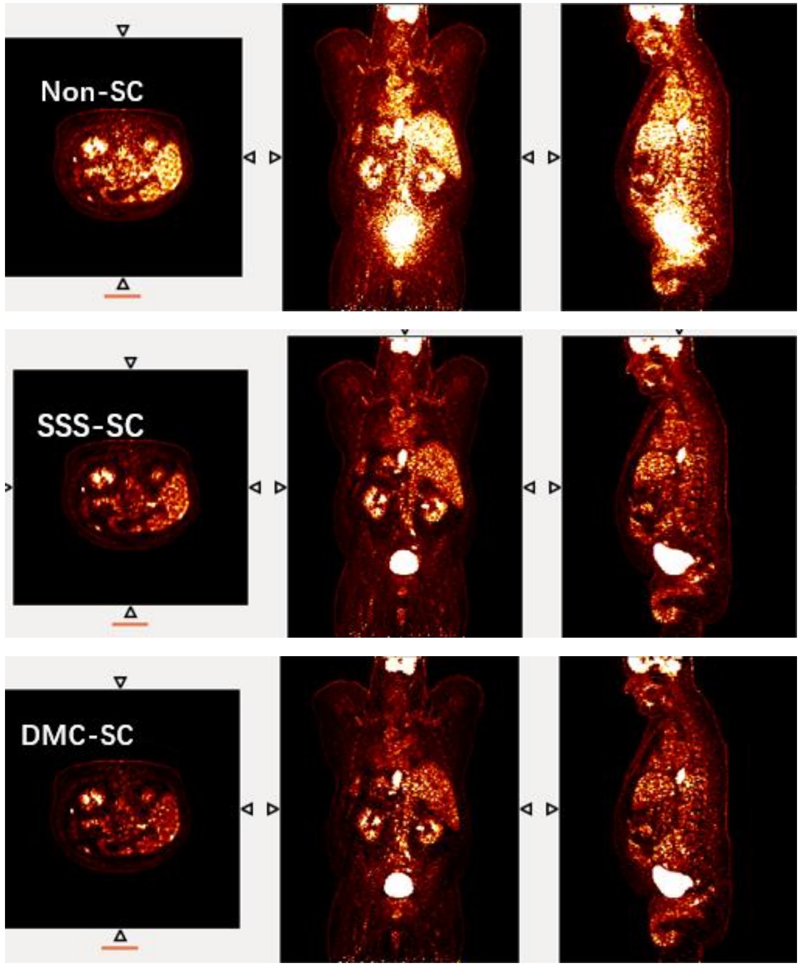
Single scatter simulation scatter correction

Neural network-based scatter correction

P1279



P2209



- Neural network based scatter correction within 10% of SSS scatter correction in liver and hotspots
- Neural network based scatter correction differs about 25% with SSS scatter correction in lungs & cold regions

Other AI Applications at CMRU

- AI-based image denoising for CT
  - ⇒ Lower dose scans
- AI-based image denoising for MR
  - ⇒ Lower field-strength scans
- AI-based image denoising for ultrasound

Together with PET AI, these form "Advanced Intelligent Clear-IQ Engine" (AiCE) for all modalities

- "AutoStroke" for image analysis and categorization
  - ⇒ Detection of signs of ischemic and hemorrhagic stroke
- Altivity for combining AI based approaches
  - ⇒ AI-based image reconstruction + workflow automation

Future Vision: Full Use of AI in PET

- Scan time per bed positions and gates requiring motion correction are automatically determined by AI

- AI-based data-driven gating for beds requiring motion correction
- AI-based scatter and randoms correction for all bed positions
- AI used to improve CT-based attenuation maps for all beds
- AI-based denoising or AI-inside-the-reconstruction approaches for image reconstruction

- AI for determination of images with high likelihood of containing lesions and possible lesion locations

Academic-Industry Partnerships

Paul Kinahan, PhD, FIEEE, FAAPM, FSNMMI, FAIBME

Vice-Chair for Research

Department of Radiology

University of Washington

Disclosures and relevant background

Current

- Co-founder of PET/X LLC
- NIH Academic-Industry Partnership grant with GE Healthcare and GE Research

Completed

- 3 NIH Academic-Industry Partnership grants
- About 12 industry-sponsored research grants or projects with six companies
- Several industry advisory boards (all unpaid)

Other

- Oversight of UW Department of Radiology industry-sponsored research grants or projects

Why participate in Academic-Industry Partnerships?

Industry

Access to viewpoint of customer base, i.e. what is needed

Access to expertise

Access to data

Test products and publicize

Co-development

License existing ideas

Leverage relationships into sales

Academia

Access to leading-edge technology

Ability to influence product development

Access to expertise

Access to research tools

Ability to interact with hardware/software at a more basic level

Funding for research

Building Extended Academic-Industry Partnerships

Partnerships can be transactional, i.e. “one and done”

Ideally, however, they are based on multiple projects over an extended period

Extended partnerships can build confidence, respect, and trust, which in turn can lead to deeper and more speculative discussions

Some amazing developments have come out of extended Academic-Industry partnerships, e.g. PET/CT and several other examples

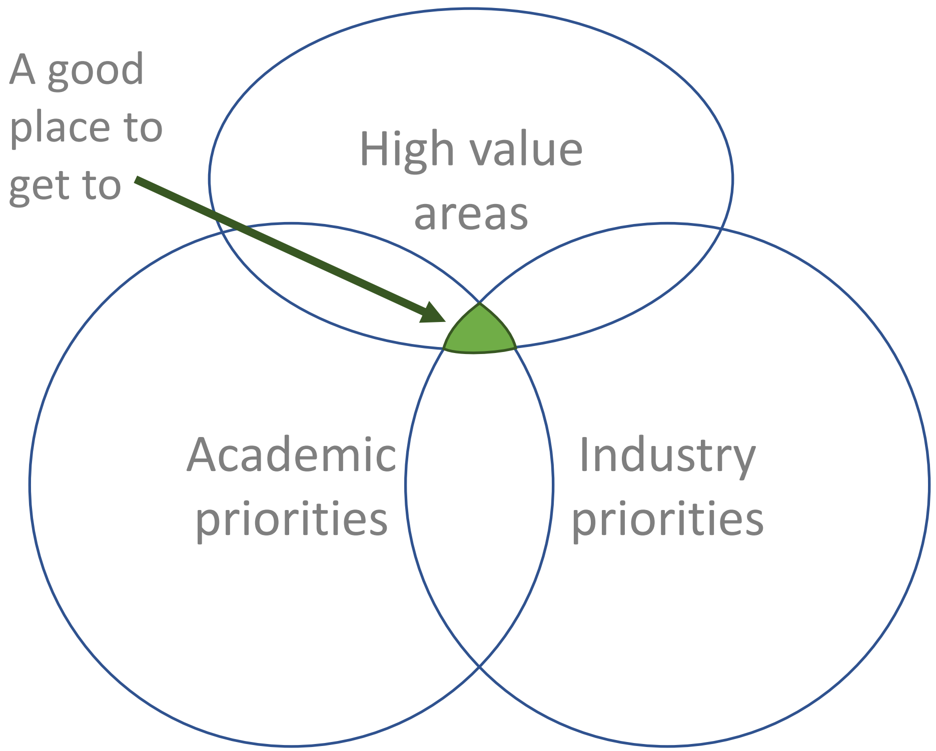
Types of research partnerships

- Data access
- Physician use, review, and publishing on products
- Unfunded collaborations
- Jointly funded projects
- Material exchange collaborations resourced by industry
  - Funding
  - Use of software or equipment or other
  - Requires "Fair market value" in exchange to be compliant, i.e. no gifts
- Licensing

Pathways to partnerships

Opening effective communications

- Start by understanding motivations and constraints in both directions
- Industry partners typically have a better understanding of motivations and constraints than academics
- Industry partners typically manage expectations more effectively
- Industry partners are typically more attentive to risk mitigation (of all kinds)
- Can often require repeated conversations and effort
- Important to stay in regular contact



Challenges in Academic-Industry Partnerships

- Time scales can be very different
- People change jobs or institution or company
- Failure to meet targets or provide deliverables
- Priorities change or key personnel have reduced time
- Delays in completing required documents (contract, COI review, IRB, DUAs etc.)

Issues for Academic-Industry Partnerships in AI

- Many newer and smaller companies
- Lots of (new) marketing that can confound understanding
- <u>Lack of curated data</u> for training, especially in molecular imaging
- Complexity of accessing data
  - Data use agreements (DUAs)
  - PHI removal from DICOM and EHR
  - Access inside hospital firewalls
- How to evaluate robustness and reproducibility
- How to test for clinical ‘fit for use’

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

NIBIB Medical Imaging and Data Resource Center (MIDRC)

A multi-group NIBIB-funded project with AAPM, ACR, and RSNA, as well as 23 other institutions

- Imaging and data commons through technology development projects
- intake portal(s) through RSNA and ACR
- imaging and data repositories/registries
- a public access portal on the Gen3 data ecosystem

Data commons enabling researchers to address topics no single archive could yield independently

for rapid and flexible collection, AI research, and dissemination of imaging and associated data

Initial research projects to expedite translation of AI from scientific findings and technical resources to public dissemination and clinical benefit

NIH National Institute of Biomedical Imaging and Bioengineering

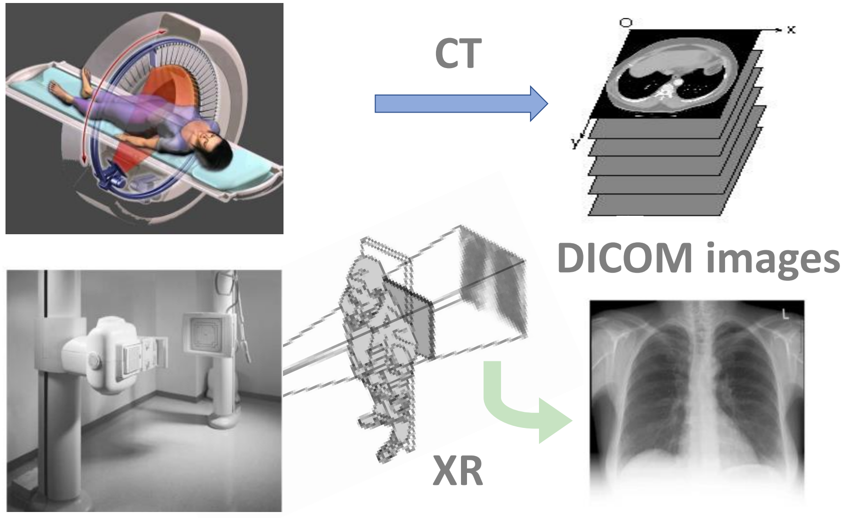
THE UNIVERSITY OF CHICAGO

AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

ACR AMERICAN COLLEGE OF RADIOLOGY

RSNA Radiological Society of North America

Many Imaging centers



Deidentification

MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

- Curation
- Annotation
- Quality Assessment
- Sequestration /Diversity
- Extraction of search data
- Presentation of search data

AI/ML algorithm developers

- Cohort selection
- Image download
- Testing
- Challenges
- Guidelines
- Metrics

Total ingested into MIDRC

\# of imaging studies

85,397

Currently undergoing MIDRC data quality and harmonization

\# of imaging studies

64,694

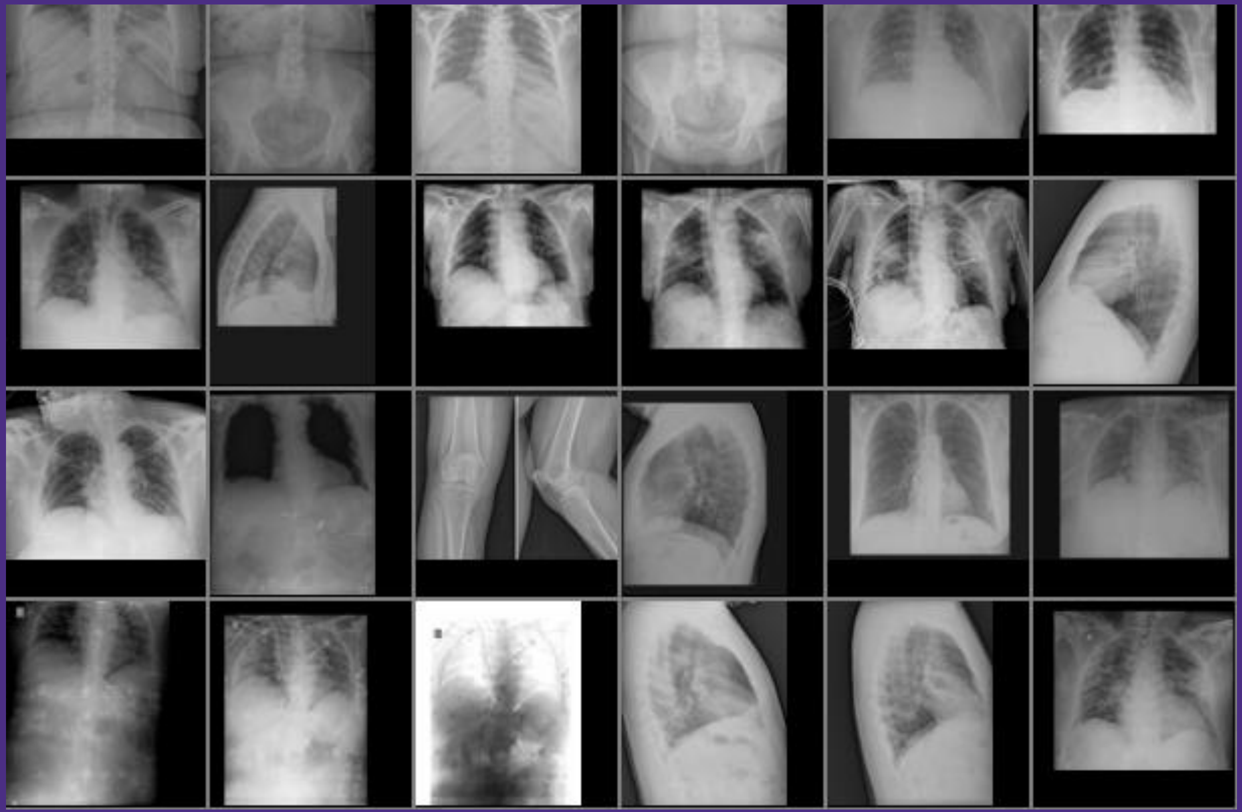
Released to the public by MIDRC

\# of imaging studies

20,703

As of 2022-02-10

The need for curation of images and DICOM meta-data



Subset of a public COVID-19 DICOM chest x-ray image collection showing variations in image quality and view directions and body part (i.e. the knee image near center)

BODY PART	STUDY DESCRIPTION	Total #	%
Head	CT HEAD WO CONTRAST	660	6%
Chest	CTA CHEST (PE STUDY) W CONTRAST	427	4%
Head	Head^HEAD (Adult)	323	3%
Abdomen	CT ABDOMEN PELVIS W CONTRAST (ROUTINE)	310	3%
Head	Vascular^PE_STUDY (Adult)	209	2%
Chest	CT CHEST WO CONTRAST	166	1%
Abdomen	Abdomen^ABD_PEL_WITH (Adult)	146	1%
Head	Head^DE_HEAD_WITHOUT_Customized (Adult)	134	1%
Chest	CT CHEST WITH CONTRAST	122	1%
Chest	CT CHEST ABDOMEN PELVIS W CONTRAST (ROUTINE	114	1%
Abdomen	CT ABDOMEN PELVIS WO CONTRAST (ROUTINE)	111	1%
Head	Head^ROUTINE_DE_HEAD (Adult)	108	1%
Abdomen	Abdomen^CT_AP_WITH (Adult)	99	1%

Section of > 350 Study Descriptions from DICOM headers for 5,500 Abdominal CT scans of patients with Covid-19, listed in order of frequency

MIDRC challenges encountered and lessons learned

- De-identification is resource intensive, and can be carried too far
- There is no national standard for description of imaging studies
- Data quality considerations include both the images and the DICOM meta-data
- We do not always need ‘high-quality’ data, rather we need data with *measured* quality
- Measuring quality for all images provided to MIDRC data not feasible
  - Wide assortment of CT and XR scanner makes and models
  - Substantial inter- and intra-center variation in imaging protocols
  - We need ‘helper AI’ for curation of large-scale data sets

Memorial Sloan Kettering Cancer Center

The clinician's needs from nuclear medicine AI

Dr. Michael J. Morris
Prostate Cancer Section Head
Member and Attending
Memorial Sloan Kettering Cancer Center

Disclosures

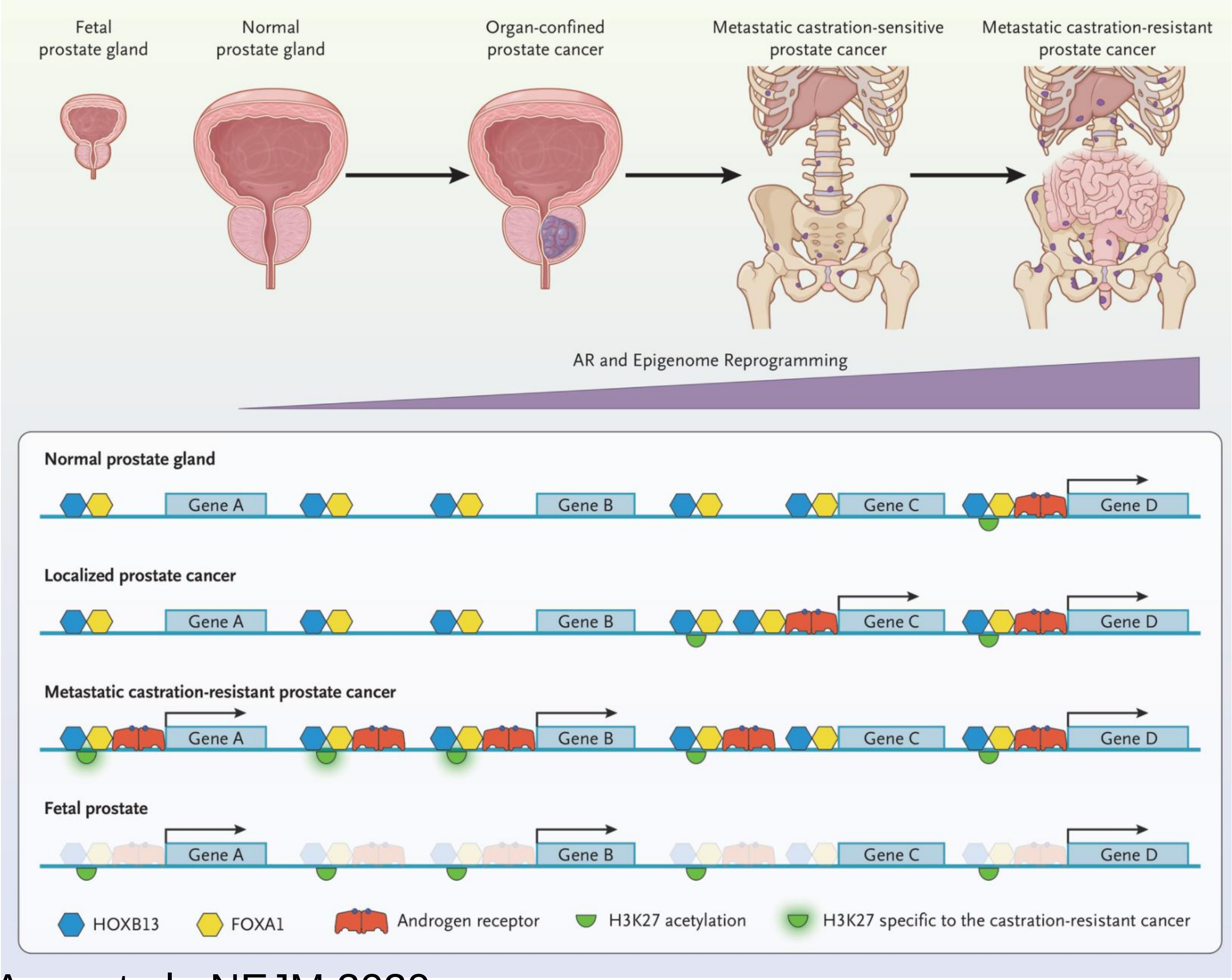
- Consultant for:
  - ORIC
  - Athenex
  - Exelixis
  - Amgen

- Institutional contracts with
  - Novartis
  - Janssen
  - Corcept
  - Celgene
  - Roche

Why do clinicians order imaging?

- Staging and treatment planning
- Prognostication
- Prediction
- Response assessments

Prostate cancer – a biologically heterogeneous disease that presents diverse clinical risks



Arap et al., NEJM 2020

The clinical quandary in prostate cancer: Is cure necessary in those for whom it is possible, and is cure possible in those for whom it is necessary?"

-- Willet Whitmore

Memorial Sloan Kettering Cancer Center

Risk assessments for localized disease

Imaging: mpMRI

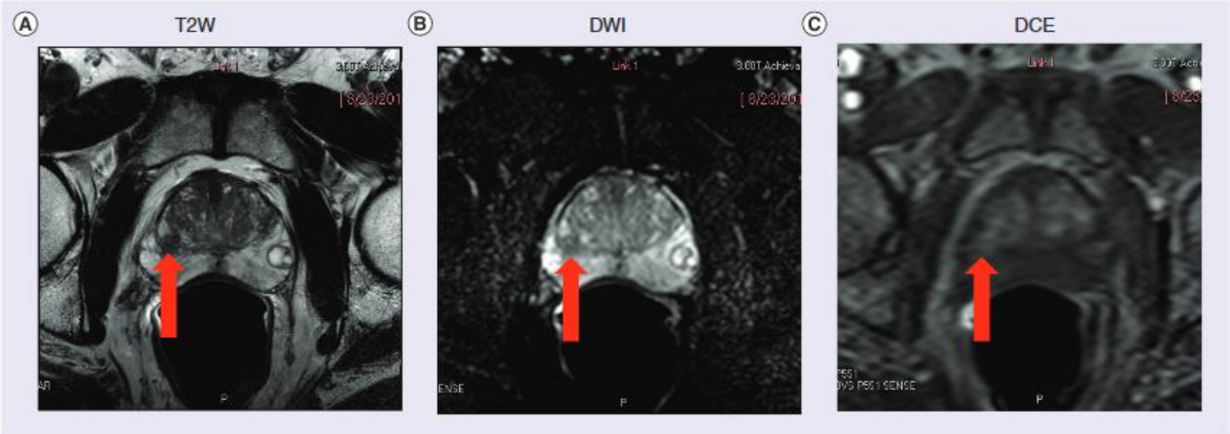


Figure 1. Multiparametric MRI of prostate from 66-year-old male with history of Gleason score 3 + 4 = 7 on systematic 12-core transrectal ultrasound-guided prostate biopsy, subsequently upgraded to Gleason score 4 + 4 = 8 following multiparametric MRI and fusion-guided targeted biopsy. Low signal intensity on T2W, restricted diffusion on DWI and focal enhancement on DCE reveal 0.7 cm lesion on right mid transitional zone (arrows).
DCE: Dynamic contrast-enhanced; DWI: Diffusion-weighted imaging; T2W: T2-weighted.

Genomic Classifiers

Test	Test material	Methodology	Sample type	Distinguishing features
Decipher	Tumor RNA expression	Whole transcriptome microarray of 22 coding and noncoding RNAs	Tissue	Predicts metastasis
Oncotype DX	Tumor RNA expression	RT-PCR of 12 cancer-related and 5 reference genes	Tissue	Predicts BCR
Prolaris	Tumor RNA expression	RT-PCR of 31 cell cycle and 15 reference genes	Tissue	Predicts BCR and metastasis
PORTOS	Tumor RNA expression	RT-PCR of 24 DNA damage, immune and radiation response genes	Tissue	Predicts response to postoperative radiation therapy
FoundationACT	Somatic mutations in cell-free DNA	ctDNA of 62 genes and 6 gene fusions	Peripheral whole blood	Advantageous if tissue is not available

Modified and adapted from Falzarano et al.[93]
BCR, biochemical recurrence; ctDNA, circulating tumor DNA; RT-PCR, reverse transcriptase polymerase chain reaction.

Serum biomarkers

Table 1. Biomarkers used to identify patients with prostate cancer.

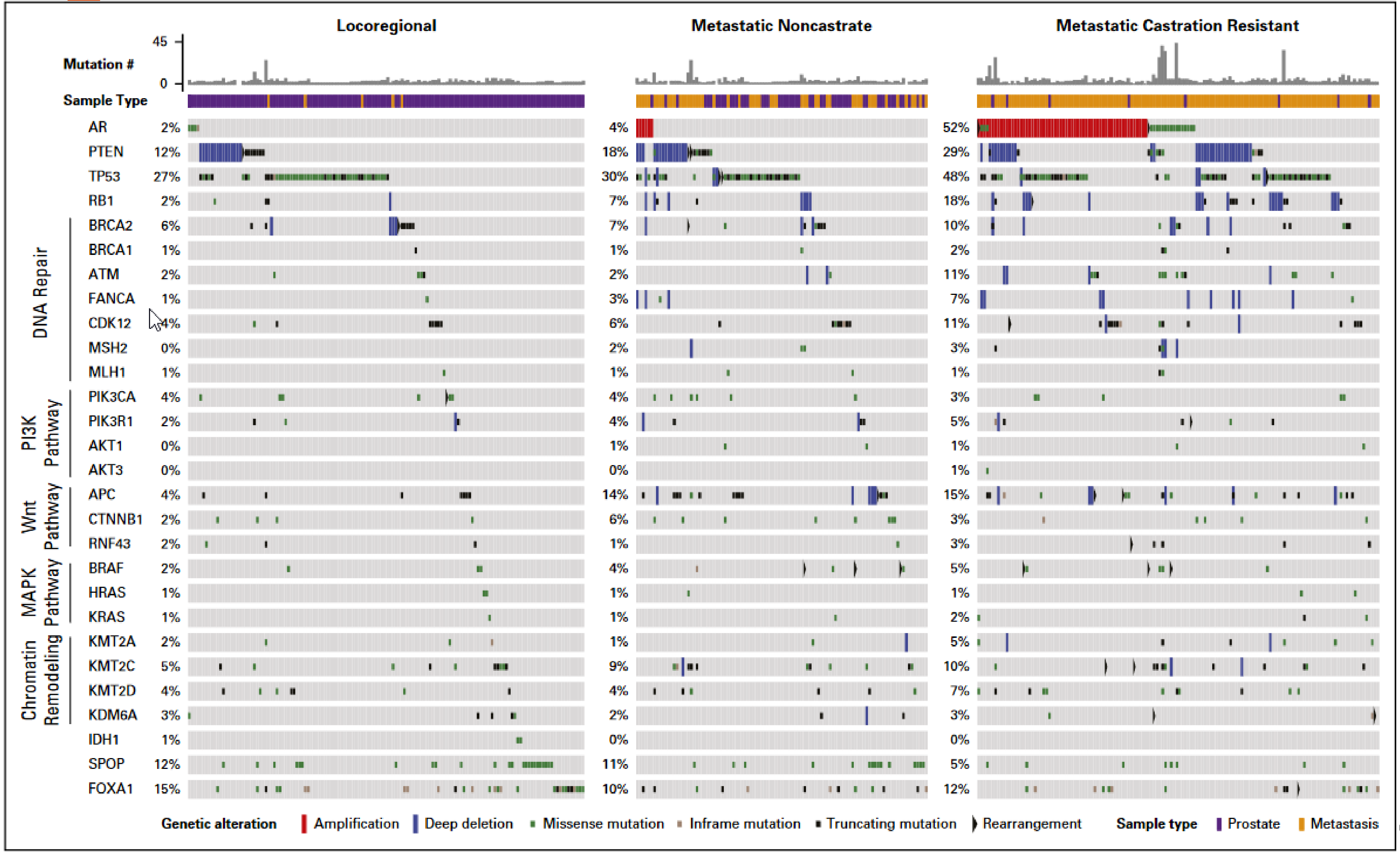
Test characteristics	Total PSA	PCA3	Prostate Health Index	4-Kallikrein score
Site of derivation	Blood	Urine	Blood	Blood
Paramater test measures	Enzyme produced exclusively by prostate cells	Overexpression of *DD3* gene (seen in 95% of PCa)	Combines three forms of PSA enzyme (total PSA, free PSA, [-2]pro-PSA)	Combines four enzymes (total PSA, free PSA, intact PSA, human KLK2)
AUC for predicting PCa	0.678	0.75	0.7	0.821

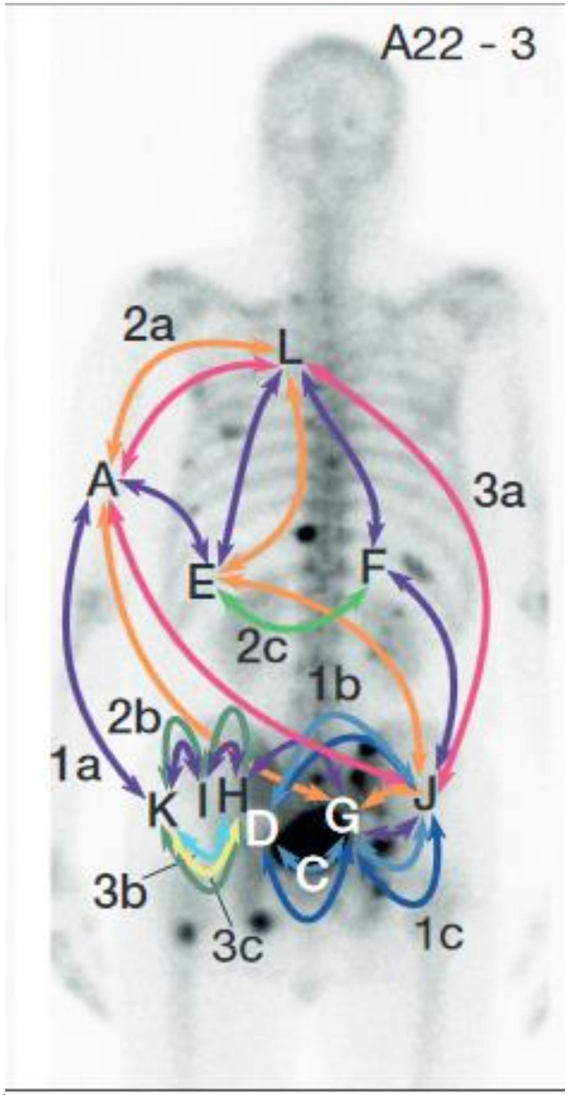
AUC: Area under receiver operator characteristic curve; PCa: Prostate cancer; PSA: Prostate-specific antigen.

Watson, Future Oncol, 2016

Memorial Sloan Kettering Cancer Center

Prostate cancer becomes increasingly biologically complex as it progresses

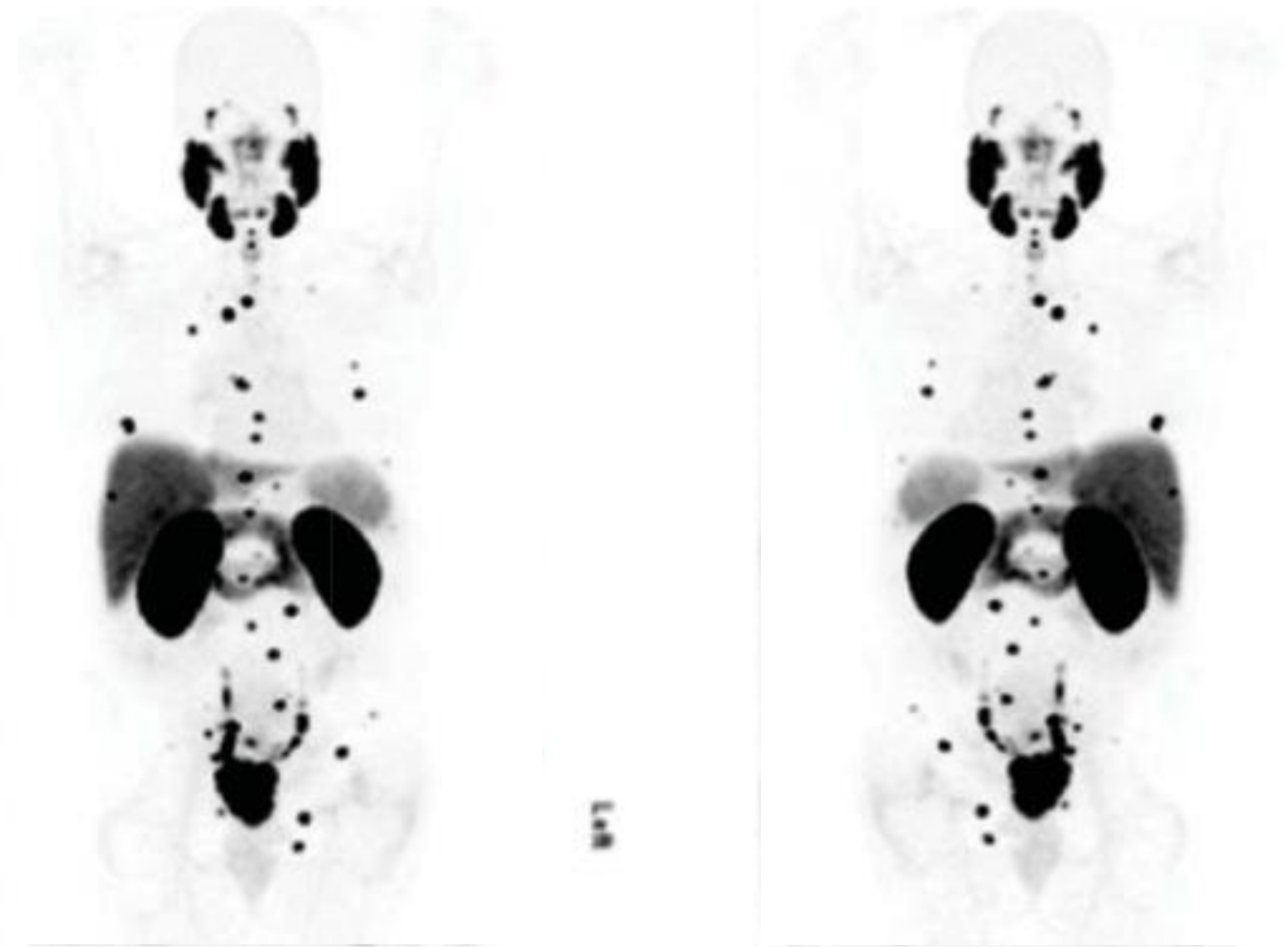




Memorial Sloan Kettering Cancer Center

Gundem et al, Nature 2015; Abida et al., JCO Precision Oncol, 2017

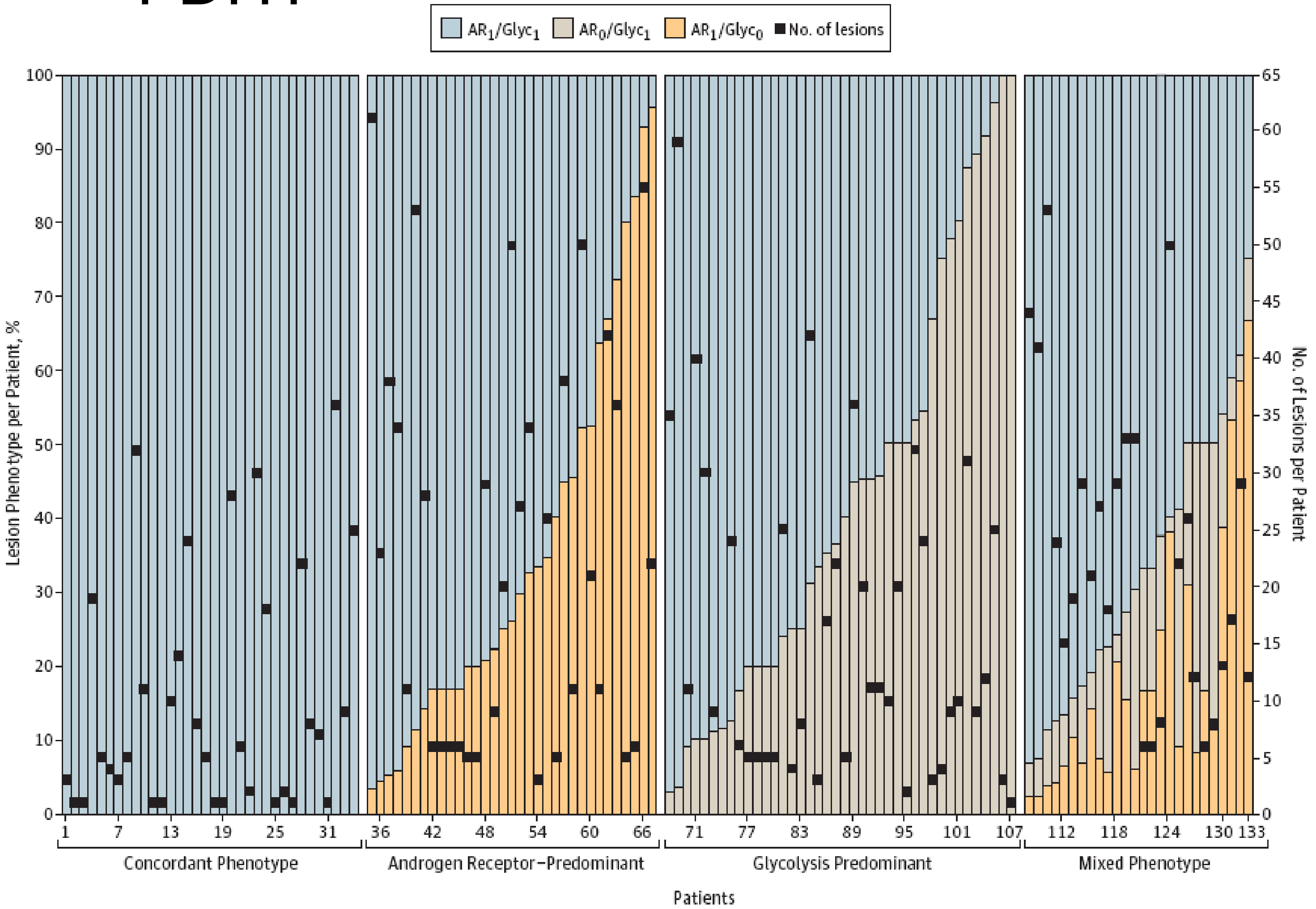
Bone scan vs. PSMA PET of the same patient



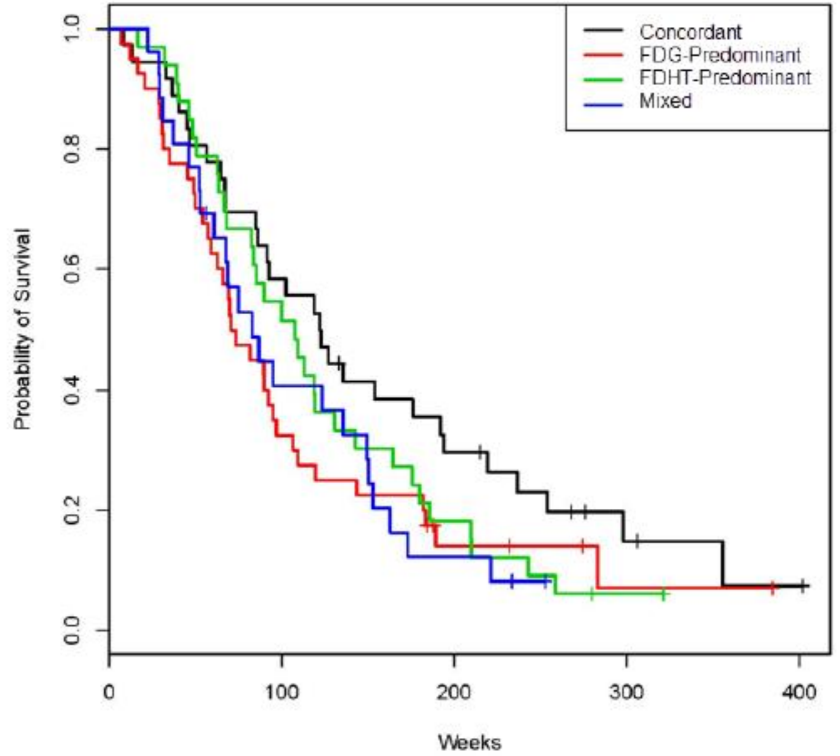
Memorial Sloan Kettering Cancer Center

Molecular imaging allows us to appreciate lesional diversity and prognosticate

PET imaging using FDG and FDHT



Likelihood of Survival



Fox et al, JAMA Oncol 2018

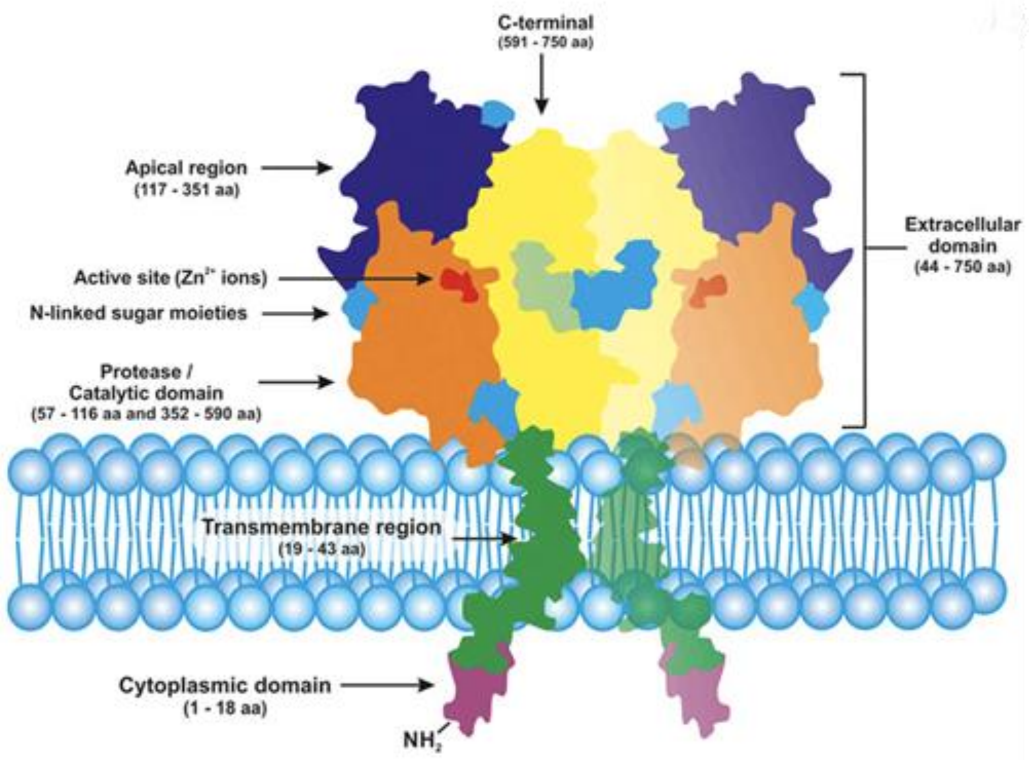
AI deliverables: staging and prognostication

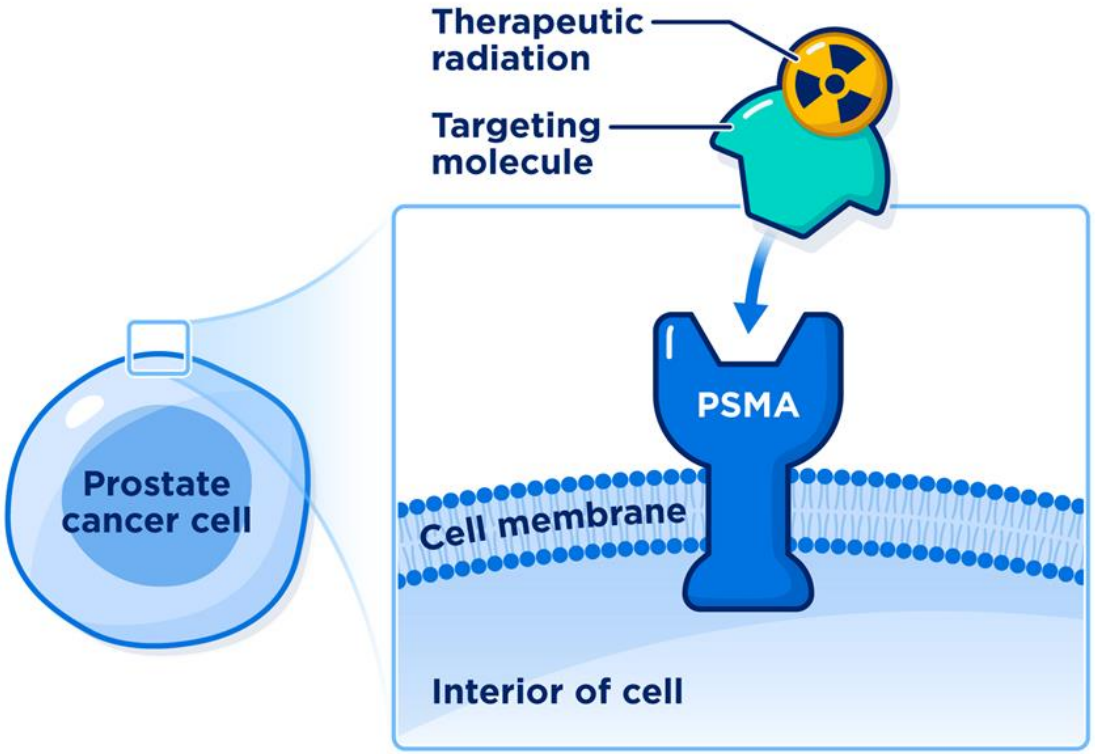
- Volume:
  - More accurate, quantifiable, clinically meaningful descriptors of disease volume

- Distribution:
  - Quantitative expressions and models of the clinical import of disease distribution (Liver > lung > bone > nodes)

- Biology:
  - Which are the lethal lesions?
  - What is the *intrapatient* and *interlesional* and *intralesional* heterogeneity, and what does that tell us about outcome?

AI and *Prediction*… key for the era of theranostics





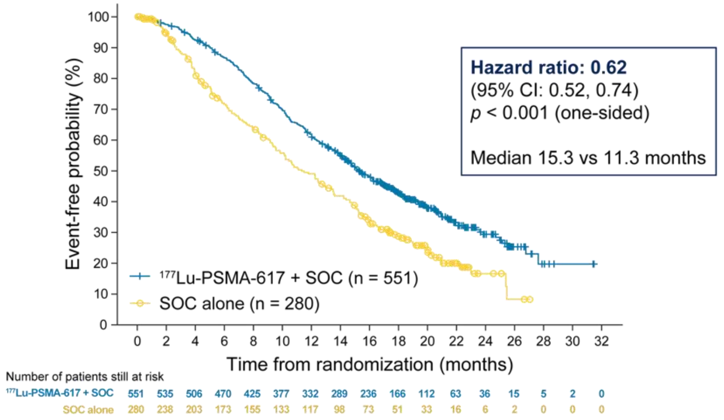
Present across disease *sites* and disease *spectrum*
Conserved in most normal tissues

O'driscoll at al Br J Pharmocol, 2016

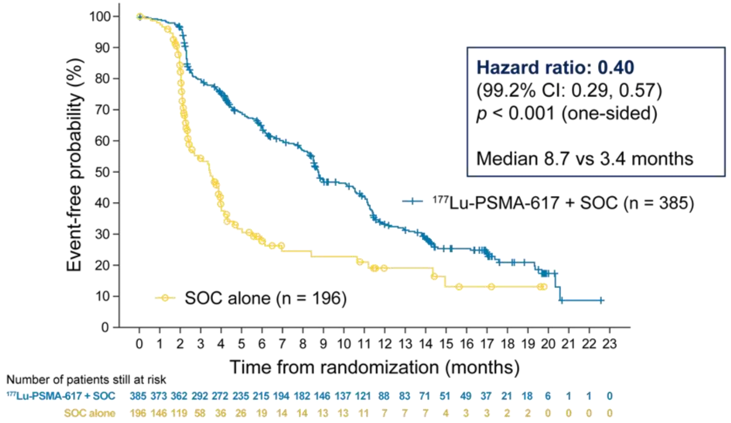
Memorial Sloan Kettering
Cancer Center

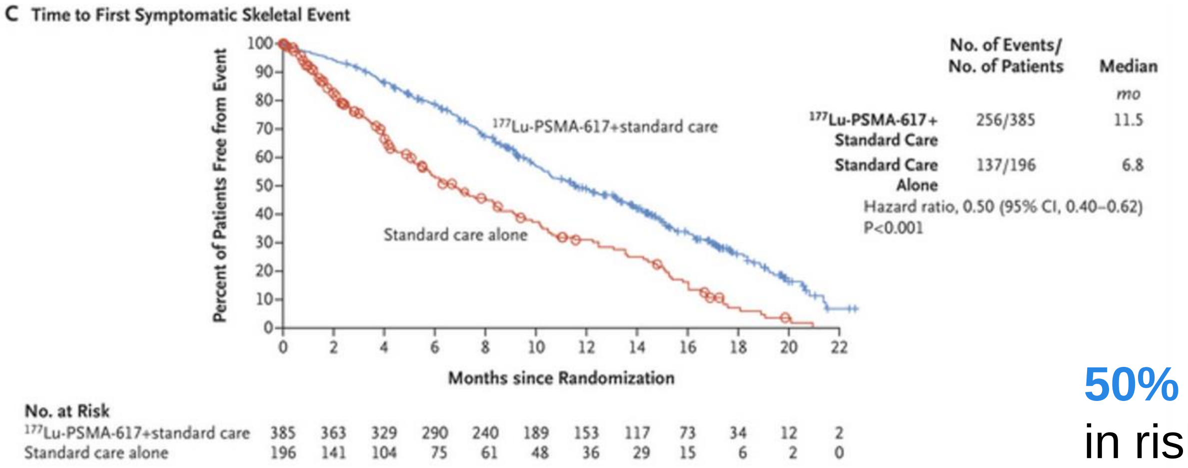
Lu-177 PSMA617 radioligand therapy prolongs life, delays progression and delays SSE's[1,2]

38% reduction in risk of death



60% reduction in risk of progression or death



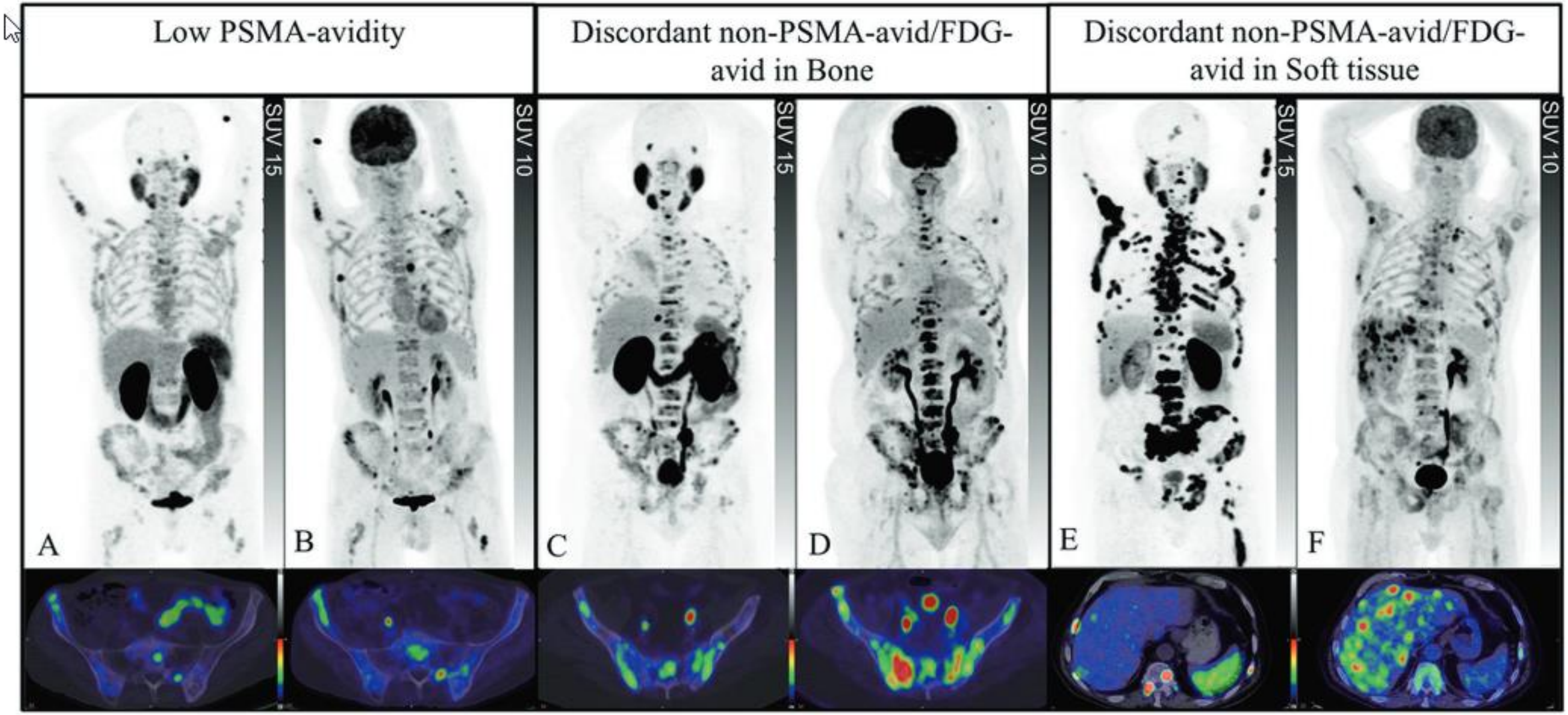


50% reduction in risk of SSE

In the Phase 3 VISION trial, 17.4 % patients were treated with Radium-223 prior to randomization in trial

1. Morris MJ, et al. J Clin Oncol. 2021;39(18_suppl):LBA4-LBA4. 2. Sartor O, et al. NEJM. 2021;(NEJMoa2107322). doi:10.1056/NEJMoa2107322.

Mechanisms of disease resistance: Heterogeneity of PSMA Expression (i.e., dose delivered)



Iravani et al., Prost Ca and Prost Dis, 2020

AI's role for predictive biomarkers

- Lesional target expression by:
  - Volume
  - Organ (bone vs. liver vs. node)

- Interaction of multiplicity of imaging modalities

- Predicted radiation dose to lesions and normal organs

- The deliverable is a model that tells us *whom* to treat, and at what dose, and then how they are responding?

Response Assessments

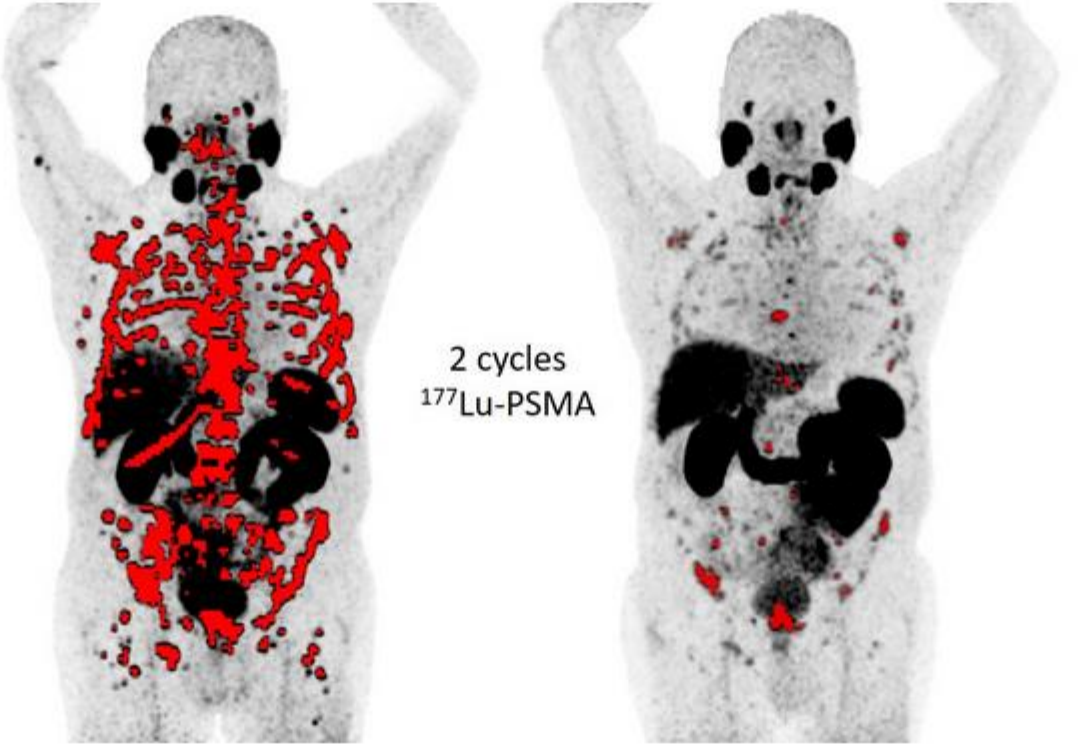
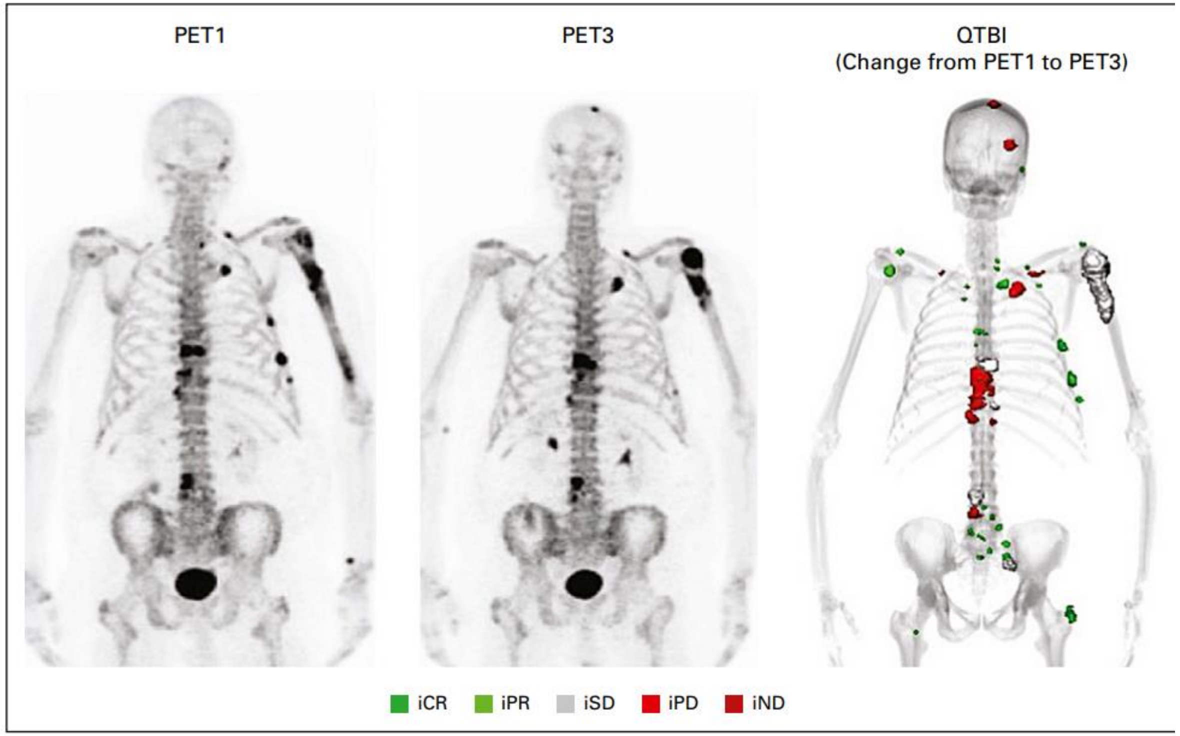
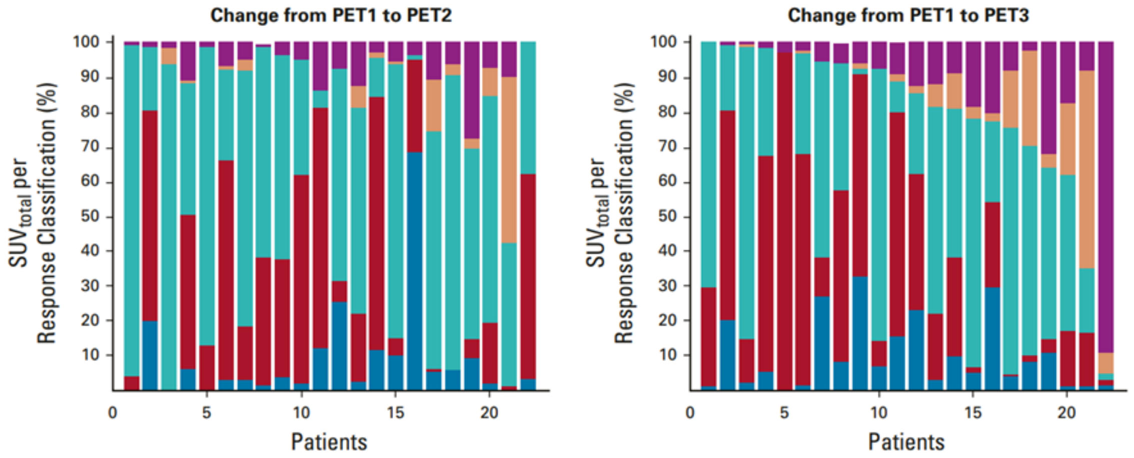


Figure: ^{68}Ga-PSMA11 PET MIP images at baseline and after 2 cycles of ^{177}Lu-PSMA radioligand therapy in a patients with mCRPC. *Image Credit: Andrei Gafita, Matthias Eiber, TUM School of Medicine, Klinikum rechts der Isar, Munich, Germany.*





iCR iPR iSD iPD iND

https://www.snmmi.org/NewsPublications/NewsDetail.aspx?ItemNumber=31986
Kyriakopoulos, JCO 2019

Memorial Sloan Kettering Cancer Center

AI's deliverable for response assessments

- Response completeness/depth
- Response kinetics
- Uniformity
- A model by which you can generate a quantitative metric, from which you can distinguish good vs. poor responses (treat vs. not treat decision)

What I Want From AI
Clinical Perspective

Eliot Siegel, MD, FSIIM, FACR
University of Maryland School of Medicine Department of Diagnostic Radiology and Nuclear Medicine
Chief Imaging VA Maryland Healthcare System

“Ground-Breaking” Filmless Department and Pandora’s Box in 1993

Any Image Any Where Any Time
Digital Enhancement and Diagnosis

Ironically, **Ground-Breaking AI is the Pandora’s Box of the 2020’s**
Exciting Promise of: Improved Accuracy, Efficiency, Safety and Information Exchange

We Still Have Not Realized the Promise of AI in 2022
What I Have from AI Today

- Detection of abnormalities
  - Lung nodules, peripheral perfusion defects on lung perfusion scan
- Diagnostic decision support
  - Probability those lung nodules are cancer?
  - Probability of PE estimated using nuclear lung scan
- Quantification
  - Measurement of lung lesions
  - SUV burden of lymphadenopathy on PET/CT
- Triage
- Segmentation

What Do I Want From AI

- Analysis over time and not for single exam which corresponds to what we actually do as nuclear medicine physicians especially for oncology applications
  - Task is evaluating change over time as often as making a new diagnosis
  - AI algorithms have been designed to plot change over time but not take change over time into consideration
  - This is one of the most critical flaws of systems today
- Customization to become optimized for a particular institution, nuclear medicine physician, region, patient population etc and to reduce bias, follow my gold standard rather than someone else's
- AI integrated with my workflow invoked dynamically when I need it, not only PACS but clinical workflow
- AI can be consumed locally as well as from the cloud
- AI that takes into account a priori probability of disease e.g. PE determination or PLCO example
- AI that makes it more efficient for me to report

- AI that increases reading efficiency and does advanced hanging protocols and generates impressions from my observations and findings
- AI that is explainable where I can intuitively understand that it is working and how
- AI that can give me its level of confidence
- AI that does population health/screening, e.g. imaging for Alzheimer's disease, maybe?
- Quality assessment AI so I can improve quality of diagnostic studies
- Natural language understanding especially new transformer natural language understanding models
  - Empathy

What Are Some Current Non Pixel Based Ground Breaking Advances in AI for Medical Imaging?

AI for Reduction in Patient "No Show" Rates: Implications for Pandemic No Show Predictions

- Chong et al demonstrated that their machine learning predictive analytics program had an AUC of 0.746 in predicting no shows resulting in a 17% reduction in the no show rate after 6 months of deployment
- Applied across the board, especially for high tech studies this could result in major improvements and importantly adaptive learning as reasons for no shows change with arrival and departure of pandemic waves of disease

AJR American Journal of Roentgenology
Diagnostic Imaging and Related Sciences
Published by ARRS Your Medical Imaging Society

You have requested the following article:

American Journal of Roentgenology, Ahead of Print : pp. 1-8

Artificial Intelligence Predictive Analytics in the Management of Outpatient MRI Appointment No-Shows

Le Roy Chong, Koh Tzan Tsai, Lee Lian Lee, Seck Guan Foo ... Show all

https://doi.org/10.2214/AJR.19.22594

ABSTRACT :

OBJECTIVE. Outpatient appointment no-shows are a common problem. Artificial intelligence predictive analytics can potentially facilitate targeted interventions to improve efficiency. We describe a quality improvement project that uses machine learning techniques to predict and reduce outpatient MRI appointment no-shows.

MATERIALS AND METHODS. Anonymized records from 32,957 outpatient MRI appointments between 2016 and 2018 were acquired for model training and validation along with a holdout test set of 1080 records from January 2019. The overall no-show rate was 17.4%. A predictive model developed with XGBoost, a decision tree-based ensemble machine learning algorithm that uses a gradient boosting framework, was deployed after various machine learning algorithms were evaluated. The simple intervention measure of using telephone call reminders for patients with the top 25% highest risk of an appointment no-show as predicted by the model was implemented over 6 months.

RESULTS. The ROC AUC for the predictive model was 0.746 with an optimized F1 score of 0.708; at this threshold, the precision and recall were 0.606 and 0.852, respectively. The AUC for the holdout test set was 0.738 with an optimized F1 score of 0.721; at this threshold, the precision and recall were 0.605 and 0.893, respectively. The no-show rate 6 months after deployment of the predictive model was 15.9% compared with 19.3% in the preceding 12-month preintervention period, corresponding to a 17.2% improvement from the baseline no-show rate ($p < 0.0001$). The no-show rates of contactable and noncontactable patients in the group at high risk of appointment no-shows as predicted by the model were 17.5% and 40.3%, respectively ($p < 0.0001$).

CONCLUSION. Machine learning predictive analytics perform moderately well in predicting complex problems involving human behavior using a modest amount of data with basic feature engineering, and they can be incorporated into routine workflow to improve health care delivery.

Keywords: artificial intelligence, machine learning, MRI, no-show, XGBoost

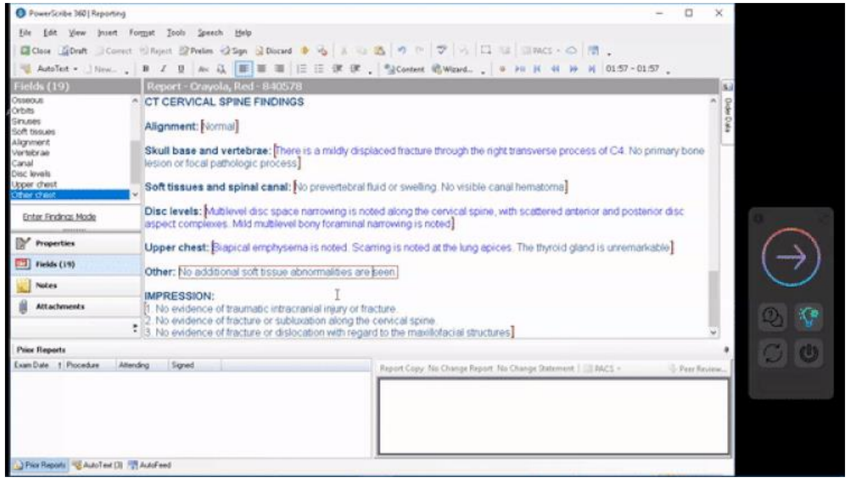
Based on a presentation at the Singapore Radiological Society 2019 annual meeting, Singapore.

Address correspondence to L. R. Chong (chong.le.roy@singhealth.com.sg).

Automation of Customized Report Impressions

- A surprising amount of time is spent dictating radiology impressions -- up to **one-third** of the entire time spent on each study, depending on modality

- AI can be used to automatically generate report impressions customized to each individual radiologist's language and preferences



- Based on initial results savings in the range of a **24%** of the total time radiologists spend on CTs -- while also decreasing radiologists' mental workload and risk of burnout

What if AI Only Read Cases Where It Was Very Confident in its Detection/Diagnosis?

Improving Workflow Efficiency for Mammography with AI Screening out Normals

- Kyono et al documented Deep Learning could achieve a 0.99 negative predictive value while excluding 34% of mammograms when there was a 15% prevalence of disease but more importantly could interpret 91% of negative mammograms when prevalence of cancer was 1%
  - Thus reducing the number of studies a mammographer would need to read by up to 91%

JACR Journal of the American College of Radiology

Log in Register Sub

ORIGINAL ARTICLE DATA SCIENCE | VOLUME 17, ISSUE 1, P56-63, JANUARY 01, 2020

Purchase Subscribe Save

Improving Workflow Efficiency for Mammography Using Machine Learning

Trent Kyono, MS • Fiona J. Gilbert, MBChB • Mihaela van der Schaar, PhD

Published: May 30, 2019 • DOI: https://doi.org/10.1016/j.jacr.2019.05.012 • Check for updates

Abstract

Key Words

References

Article Info

Related Articles

Abstract

Objective

The aim of this study was to determine whether machine learning could reduce the number of mammograms the radiologist must read by using a machine-learning classifier to correctly identify normal mammograms and to select the uncertain and abnormal examinations for radiological interpretation.

Methods

Mammograms in a research data set from over 7,000 women who were recalled for assessment at six UK National Health Service Breast Screening Program centers were used. A convolutional neural network in conjunction with multitask learning was used to extract imaging features from mammograms that mimic the radiological assessment provided by a radiologist, the patient's nonimaging features, and pathology outcomes. A deep neural network was then used to concatenate and fuse multiple mammogram views to predict both a diagnosis and a recommendation of whether or not additional radiological assessment was needed.

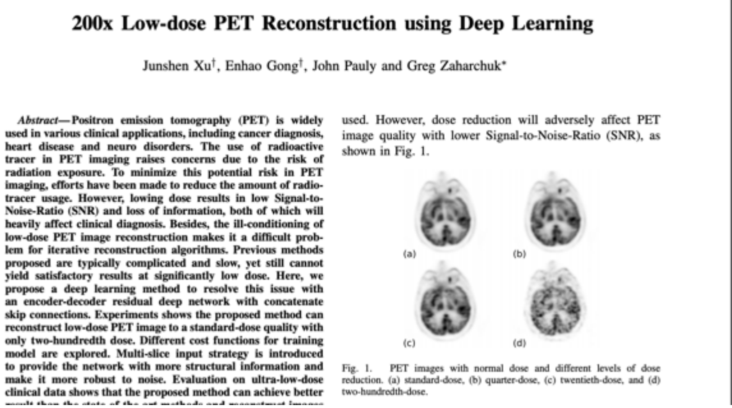
Results

Ten-fold cross-validation was used on 2,000 randomly selected patients from the data set; the remainder of the data set was used for convolutional neural network training. While maintaining an acceptable negative predictive value of 0.99, the proposed model was able to identify 34% (95% confidence interval, 25%-43%) and 91% (95% confidence interval;

What Will Be the Initial "Killer App" for AI (Deep Learning) in Diagnostic Imaging?

- So, it turns out that we can not only use Deep Learning to detect and diagnose and quantify, but we can also create images using AI
- Immediate benefits from ubiquitous adoption by manufacturers of Deep Learning for Image Acquisition and processing
  - Major MRI and CT and nuclear medicine vendors will soon adopt Deep Learning to substantially improve image quality, especially texture and reduce scan times and doses
  - Iterative reconstruction sacrifices texture for reduced noise but Deep Learning can optimize image quality without reduction in important diagnostic features
  - Model based iterative reconstruction optimizes trade-offs but is highly computationally intensive and this has been a major limiting step in its use in day to day scanning

AI Has and Will Revolutionize Image Acquisition in Diagnostic Imaging



200x Low-dose PET Reconstruction using Deep Learning

Junshen Xu†, Enhao Gong†, John Pauly and Greg Zaharchuk*

***Abstract*—Positron emission tomography (PET) is widely used in various clinical applications, including cancer diagnosis, heart disease and neuro disorders. The use of radioactive tracer in PET imaging raises concerns due to the risk of radiation exposure. To minimize this potential risk in PET imaging, efforts have been made to reduce the amount of radio-tracer usage. However, lowing dose results in low Signal-to-Noise-Ratio (SNR) and loss of information, both of which will heavily affect clinical diagnosis. Besides, the ill-conditioning of low-dose PET image reconstruction makes it a difficult problem for iterative reconstruction algorithms. Previous methods proposed are typically complicated and slow, yet still cannot yield satisfactory results at significantly low dose. Here, we propose a deep learning method to resolve this issue with an encoder-decoder residual deep network with concatenate skip connections. Experiments shows the proposed method can reconstruct low-dose PET image to a standard-dose quality with only two-hundredth dose. Different cost functions for training model are explored. Multi-slice input strategy is introduced to provide the network with more structural information and make it more robust to noise. Evaluation on ultra-low-dose clinical data shows that the proposed method can achieve better**

used. However, dose reduction will adversely affect PET image quality with lower Signal-to-Noise-Ratio (SNR), as shown in Fig. 1.

(a) (b)

(c) (d)

Fig. 1. PET images with normal dose and different levels of dose reduction. (a) standard-dose, (b) quarter-dose, (c) twentieth-dose, and (d) two-hundredth-dose.

- Pandemic will increase pressure to scan patients more rapidly with reductions in scanning time decreasing patient and staff exposure and improving efficiency
- AI, "Deep Learning" for reconstruction of CT, MRI, PET, conventional radiography will become the "killer app" of 2022

AI Towards Precision Medicine
Selective/Smart Screening

- Screening can be smarter by more precisely identifying populations at risk for certain diseases
- This will decrease the number of patients that need to be screened while increasing the yield of screening for disease

A Priori Probability of Disease: PLCO

- Published in 2009, the PLCO Screening Trial enrolled ~155,000 participants to determine whether certain screening exams reduced mortality from prostate, lung, colorectal and ovarian cancer
- The Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial dataset provides an unparalleled resource for matching patients with the outcomes of demographically or diagnostically comparable patients
- These matched data can be used to inform a more sophisticated, personalized diagnostic decision-making process by tailoring imaging and testing follow-up intervals or even guiding intervention and prognosis
- They can also be incorporated into CAD algorithms to improve diagnostic efficacy by provided a priori likelihood of disease information.

PLCO Dataset

Additional Criteria: African American, Native Hawaiian or Pacific Islander, Family or Personal History Cancer, COPD

Table 2. Modified Logistic-Regression Prediction Model ($PLCO_{M2012}$) of Cancer Risk for 36,286 Control Participants Who Had Ever Smoked.*

Variable	Odds Ratio (95% CI)	P Value	Beta Coefficient
Age, per 1–yr increase†	1.081 (1.057–1.105)	<0.001	0.0778868
Race or ethnic group‡			
White	1.000		Reference group
Black	1.484 (1.083–2.033)	0.01	0.3944778
Hispanic	0.475 (0.195–1.160)	0.10	−0.7434744
Asian	0.627 (0.332–1.185)	0.15	−0.466585
American Indian or Alaskan Native	1		0
Native Hawaiian or Pacific Islander	2.793 (0.992–7.862)	0.05	1.027152
Education, per increase of 1 level†§	0.922 (0.874–0.972)	0.003	−0.0812744
Body-mass index, per 1-unit increase†	0.973 (0.955–0.991)	0.003	−0.0274194
Chronic obstructive pulmonary disease (yes vs. no)	1.427 (1.162–1.751)	0.001	0.3553063
Personal history of cancer (yes vs. no)	1.582 (1.172–2.128)	0.003	0.4589971
Family history of lung cancer (yes vs. no)	1.799 (1.471–2.200)	<0.001	0.587185
Smoking status (current vs. former)	1.297 (1.047–1.605)	0.02	0.2597431
Smoking intensity¶			−1.822606
Duration of smoking, per 1-yr increase†	1.032 (1.014–1.051)	0.001	0.0317321
Smoking quit time, per 1-yr increase†	0.970 (0.950–0.990)	0.003	−0.0308572
Model constant			−4.532506

* To calculate the 6-year probability of lung cancer in an individual person with the use of categorical variables, multiply the variable or the level beta coefficient of the variable by 1 if the factor is present and by 0 if it is absent. For continuous

PLCO Participants Who Qualify for NLST: Smokers 55 to 74 Years Old

PLCO Participants Who Qualify for the National Lung Screening Trial:

All | None

- [ ] No
- [x] Yes

Age at BQ: 49 - 78

Height (inches): 48 - 84

Weight (lbs) at Baseline: 70 - 399

Gender:

All | None

- [x] Male
- [x] Female

Cigarette Smoking Status:

All | None

- [x] Never Smoked Cigarettes
- [x] Current Cigarette Smoker
- [x] Former Cigarette Smoker

Ever Smoked Cigars?:

All | None

Analyze

Results returned in 1.033 seconds

Print results

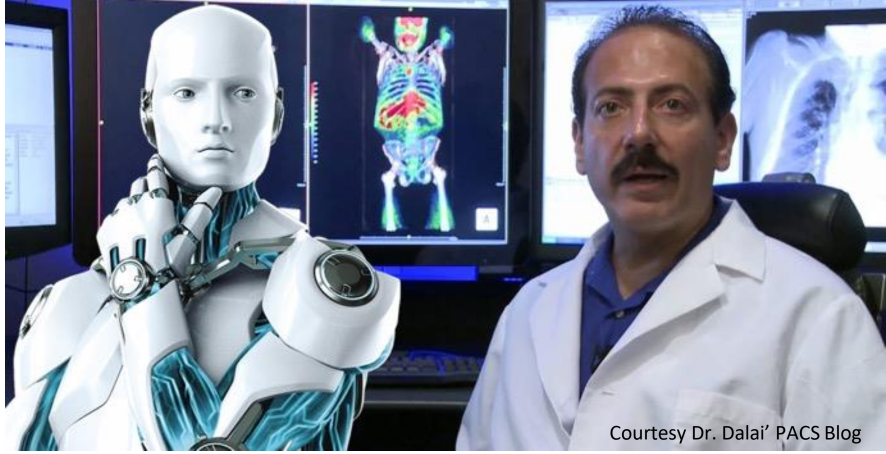
Permanent link

Total Matches (experimental): 29719

Total Matches (overall): 114697

Cancer Type	Relative Risk (95% CI)	p value	Experimental Rate (cases/total)	Control Rate (cases/total)
Mortality	2.09 (2.04-2.14)	<0.0001	24.19% (7188/29719)	11.59% (13289/114697)
All Cancers by participant	1.42 (1.39-1.45)	<0.0001	23.73% (7052/29719)	16.69% (19139/114697)
Lung	7.57 (7.05-8.12)	<0.0001	7.42% (2204/29719)	0.98% (1124/114697)
Prostate	0.8 (0.76-0.84)	<0.0001	9.37% (1690/18031)	11.78% (6266/53201)
Breast	0.73 (0.67-0.79)	<0.0001	2.32% (688/29719)	3.16% (3622/114697)
Bladder	2.54 (2.28-2.83)	<0.0001	1.82% (540/29719)	0.72% (821/114697)
Colorectum	1.28 (1.16-1.41)	<0.0001	1.8% (536/29719)	1.41% (1616/114697)
NonHodgkin's Lymphoma	0.83 (0.72-0.96)	0.0111	0.71% (210/29719)	0.85% (980/114697)
Melanoma	0.71 (0.61-0.82)	<0.0001	0.71% (210/29719)	0.99% (1[illegible]697)

Beyond Automated Image Interpretation

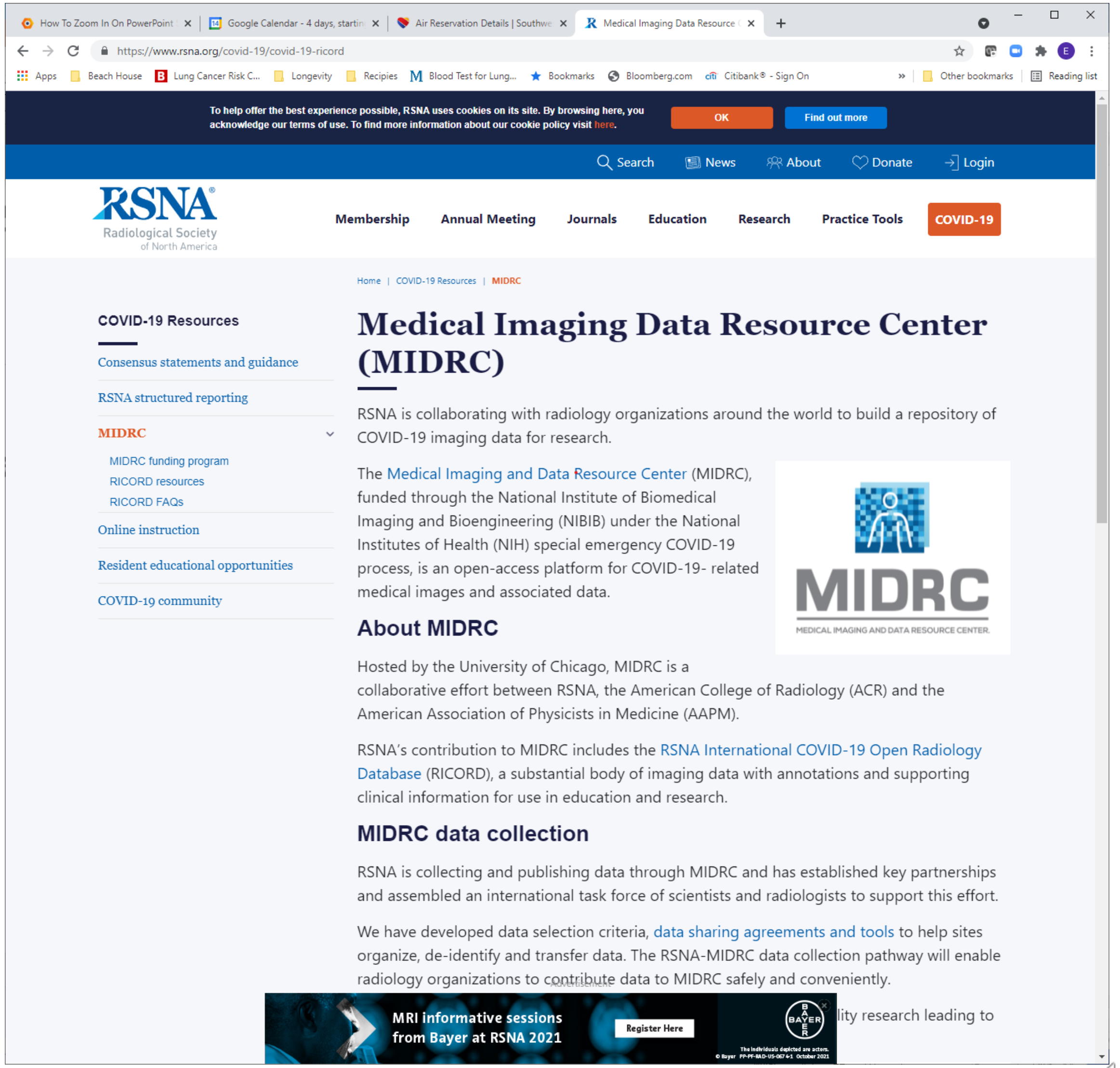


Courtesy Dr. Dalai' PACS Blog

- Much of the AI literature has been devoted to improved image interpretation
  - Lung nodules, pediatric bone age, intracranial hemorrhage, fracture detection
- Many AI Algorithms focus on quantitative assessment or on image analytics and quantification such as radiomics
  - Characterizing morphology or texture to predict tumor type, histology, grade, prognosis, etc.
- Increasingly AI algorithms will focus on enhanced efficiency/productivity in addition to enhanced accuracy/decision support

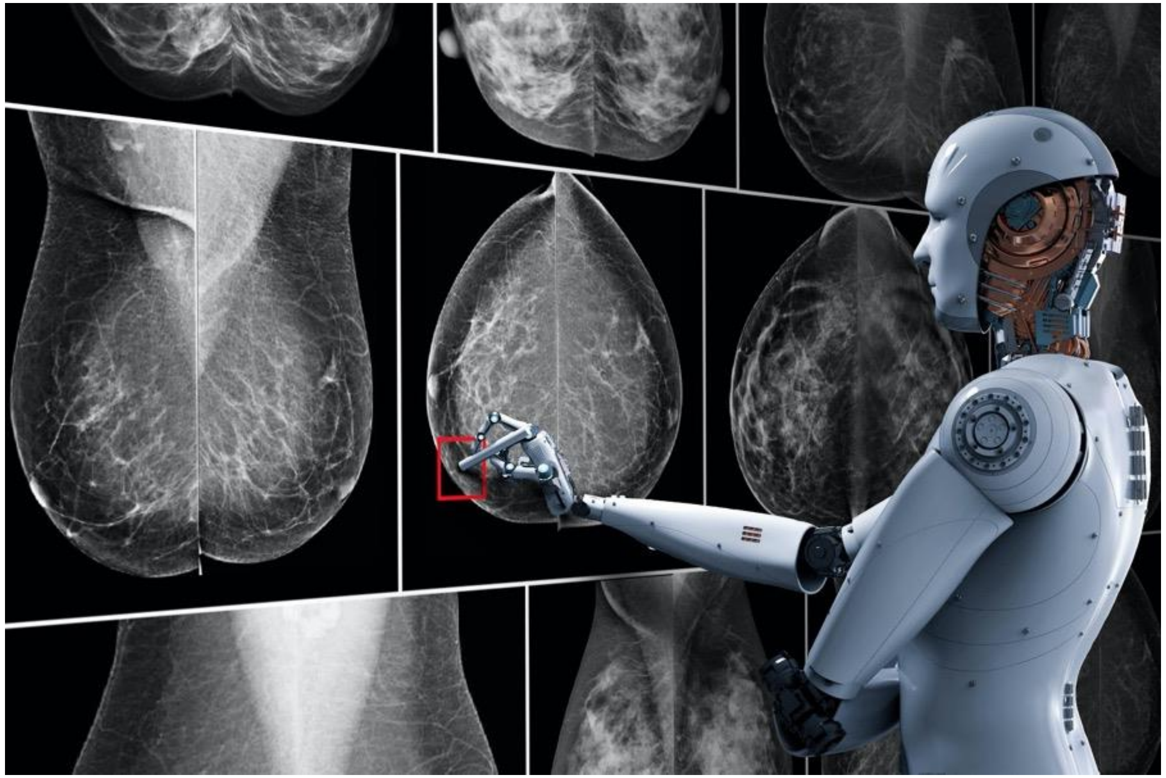
Examples Deep Learning Based on Availability of Academic Test Sets

- COVID
- Bone age
- Lung nodules – LIDC, RIDER, NLST



Examples Algorithms Developed Based on Clinical Need

- PET/CT adenopathy, marrow evaluation, renal function, cardiac and brain uptake
- Perfusion scanning
- Automated renal flow analysis
- Fractional flow reserve analysis



SIIM CMIMI20

Shaping the Future with AI

Co-sponsored by the AAPM

U.S. FOOD & DRUG ADMINISTRATION

Expanding Functionality of Machine Learning in Medical Imaging: An FDA perspective

Monday, Sep 14 | 5:25 pm – 6:15 pm ET

Berkman Sahiner, PhD

Matthew Diamond, MD, PhD

Jennifer Segui, MS

Shahram Vaezy, PhD

Nicholas Petrick, PhD

Center for Devices and Radiological Health
U.S. Food and Drug Administration

Question

- Are there ways to reasonably ensure that an AI-based software as a medical device (SaMD) remains safe and effective as the device learns, while the device sponsor does not have to come back to the FDA for particular types of modifications?

Spectrum of Modifications to ML/AI-based Algorithms

Locked Algorithm, Discrete Updates Controlled by Human Intervention

More human control

Less human control

Adaptive Algorithm, Automated Process for Updates

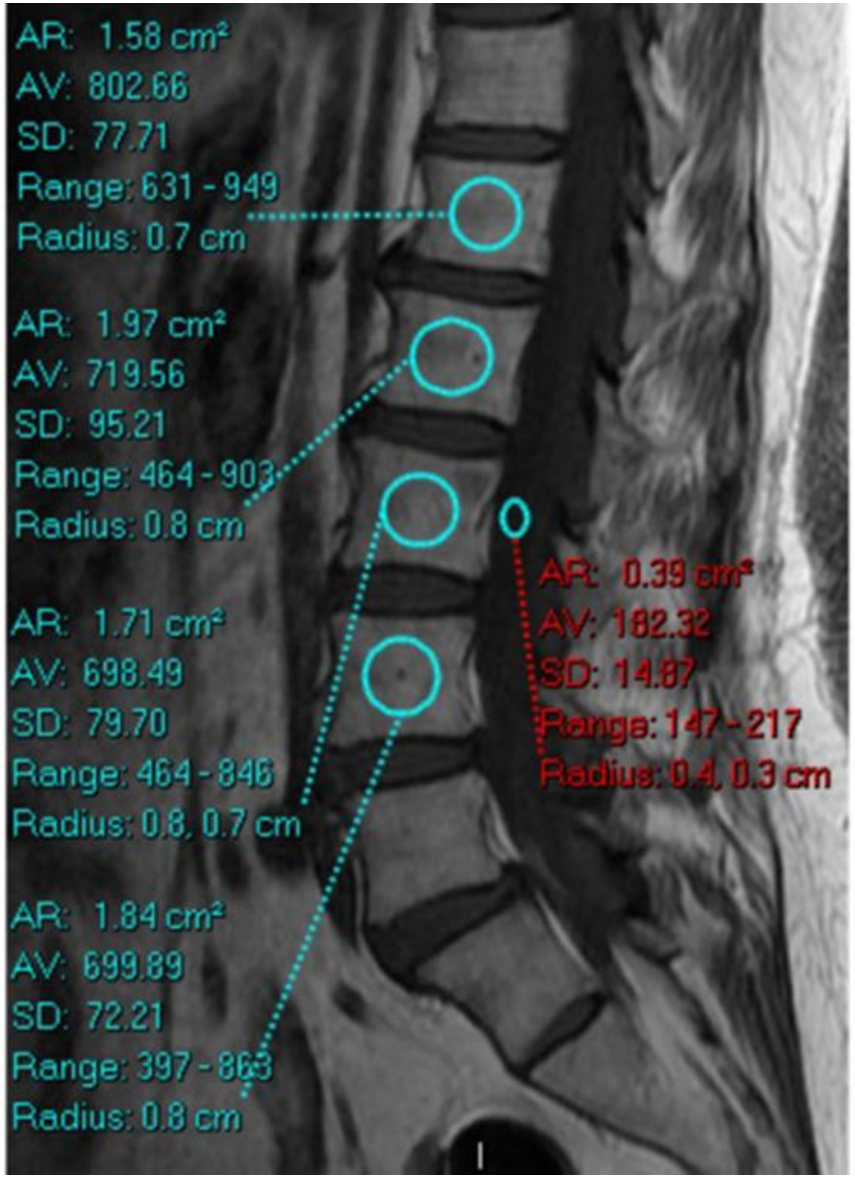
**If adaptations are pre-specified,
and the methods for determining an appropriate adaptation clearly delineated,
then the same decision-making framework may be similarly applied for both locked and adaptive algorithms**

Types of Changes: Change Related to

- Performance
  - Re-training with additional data sets within the intended use population from the same type of input signal
  - Change in the AI/ML architecture
- Inputs
  - Expanding compatibility with other source(s) of the same input data type
  - Adding different input data type(s)
- Intended use
  - Expanding the intended use population
  - Expanding use of image analysis algorithm to a different organ

Current and Potential Population Health Apply for all Patients

- Bone mineral density
  - Vertebral body fracture
- Coronary artery calcification
- Abdominal aortic aneurysm detection
- Gallstones
- Renal calculi



Vertebral bone quality score predicts fragility fractures independently of bone mineral density. Ehresman et al

Image Quality Control

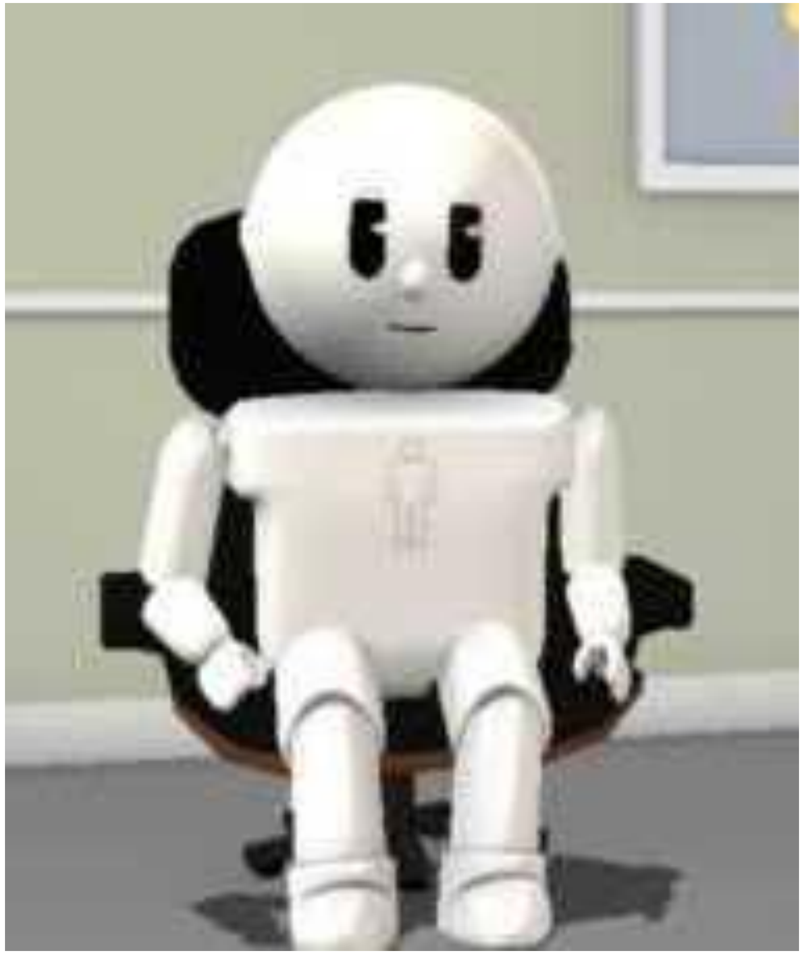
- Cardiac imaging
- PET uptake quality, evaluation of image fusion
- CT dose vs. noise quality control
- PSMA imaging quantification
- Brain PET analysis ahd quantification

What Do We Need from Next Generation AI Clinically?

- Improve efficiency/productivity
  - What's it like to practice nuclear medicine nowadays?
  - Our previous research has suggested nuclear medicicne physicians spend 85% of their time on clerical/admin/repetitive tasks and only 15% on image interpretation!
- Radiology scribes are being used by some practices to increase that 15% to beyond 50% resulting in major improvements in reading times

What We Really Want is Some Empathy

- According to a study done by the Mayo Clinic in 2006, the most important characteristics patients feel a good doctor must possess are entirely human
- According to the study, the ideal physician is confident, empathetic, humane, personal, forthright, respectful, and thorough
- Watson may have proved his cognitive superiority, but can a computer ever be taught these human attributes needed to negotiate through patient fear, anxiety, and confusion? Could such a computer ever come across as sincere?



UNIVERSITY of MARYLAND
SCHOOL OF MEDICINE

Conclusion
AI From Groundbreaking to Invisible?

- AI will undoubtedly have a major positive impact on efficiency, accuracy, discoverability, safety, and efficacy in diagnostic imaging, which will revolutionize the practice of nuclear medicine over the next decade
- This will allow our specialty to stay relevant and indeed critical as we enter the dawn of the era of personalized/precision medicine

Conclusion
AI From Groundbreaking to Invisible?

- The transition from film to digital imaging almost 30 years ago brought about not only ubiquitous access to images but also the tantalizing promise that “artificial intelligence” could be utilized for these digital images to achieve earlier and more accurate detection, diagnosis, and treatment
- 30 years later, however we are just beginning to realize the groundbreaking potential of AI

What I Want From AI
Clinical Perspective

Eliot Siegel, MD, FSIIM, FACR
Professor and Vice Chair University of Maryland School of Medicine Department of Diagnostic Radiology and Nuclear Medicine
Chief Radiology and Nuclear Medicine VA Maryland Healthcare System
Professor Computer Science UMBC
Professor Biomedical Engineering UMCP

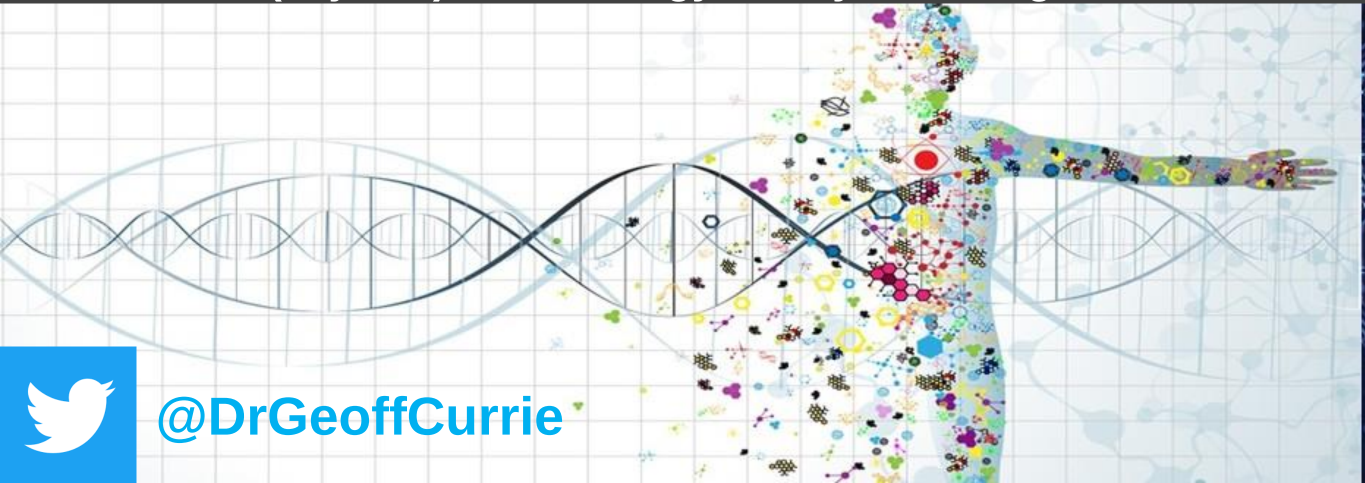
Baylor College of Medicine

Charles Sturt University

SNMMI Artificial Intelligence Summit 2022: What Do Nuclear Medicine Technologists

Dr Geoff Currie, AM

CNMT, BPharm, MMedRadSc, MAppMngt, MBA, PhD
Professor in Nuclear Medicine at Charles Sturt University
Professor (adjunct) in Radiology at Baylor College of Medicine



@DrGeoffCurrie

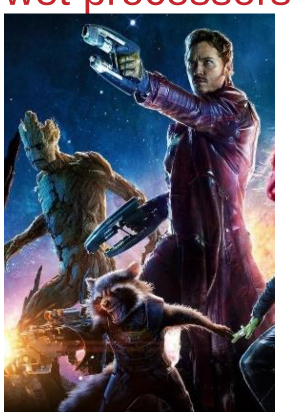
AI and the NMT!

Science denial

Not while ever the world is flat!

Guardians of the Galaxy

Not on my watch! I am still fighting to keep wet processors.



Conspiracy

Not giving in to 'the man' so they can control us or make more money!

Ostrich

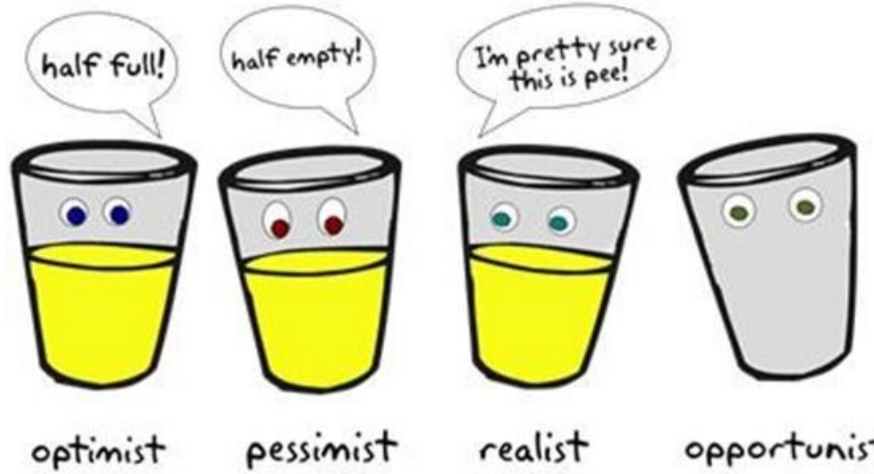
That's for the doctors and physicists to worry about. Head in the sand!

Optimist

The efficiencies will allow me to spend more time providing quality patient care!

Pessimist

I'll never be able to keep up with technology! I'll be redundant.



Realist

A lot of work to do but there are opportunities to improve outcomes if we manage potential risks!

Opportunist

There will be opportunities to improve workflows and diversify of our responsibilities! I want to get ahead of the curve.

Opportunity / Inclusion

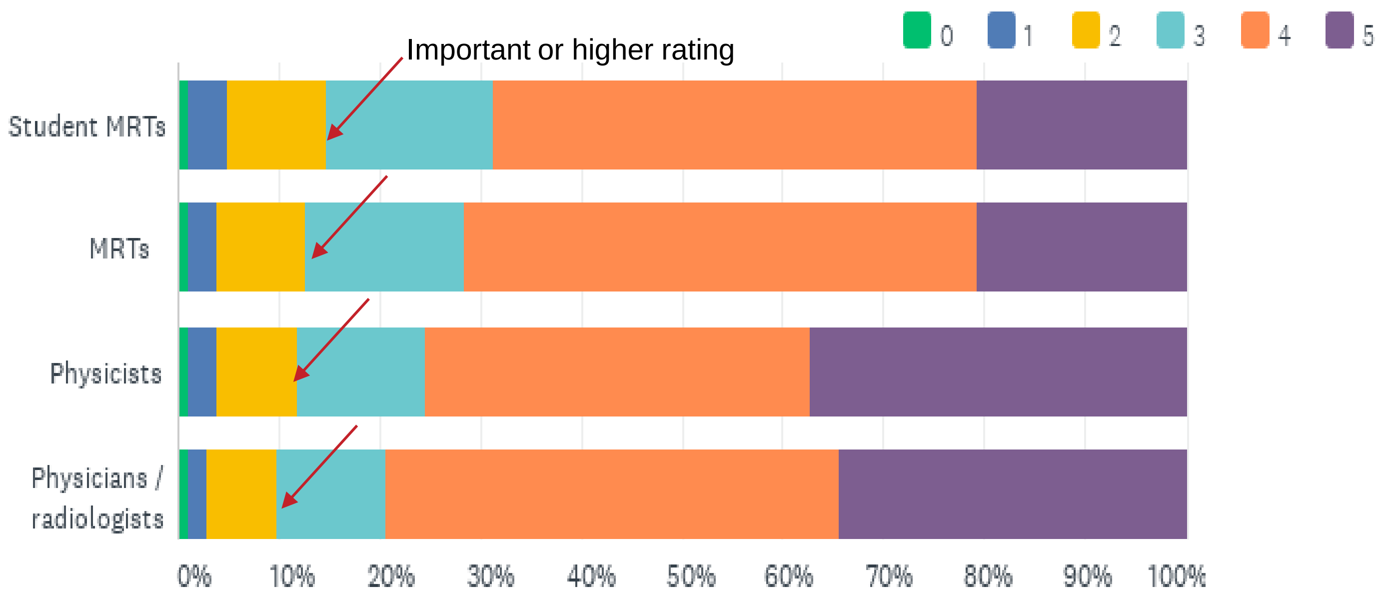
- Not all NMTs will be interested or have something to offer, but there are some pretty switched on NMTs that have a lot to offer this space. **Inclusivity = diversity**
- NMTs are front line for **clinical application** of commercialised algorithms as is current practice with SaMD (software as medical device).
- Data and **information management** as per PACS and RIS systems.
- **Image manipulation** and analysis (will implement as GUI) is often NMT driven. What are the insights of the users?
- Validation of algorithms as part of a **research** team for those working in that research domain (very few).
- Possible role extension for **data curation** / stewardship.

NMT Needs from AI

- Representativeness and voice
- Nothing about us, without us!
- Avoid imposing technology and change
- Recognise as stakeholder
- Common language
- Transparency of application
- Inclusivity
- Education and understanding
- Awareness
- Security of position and role

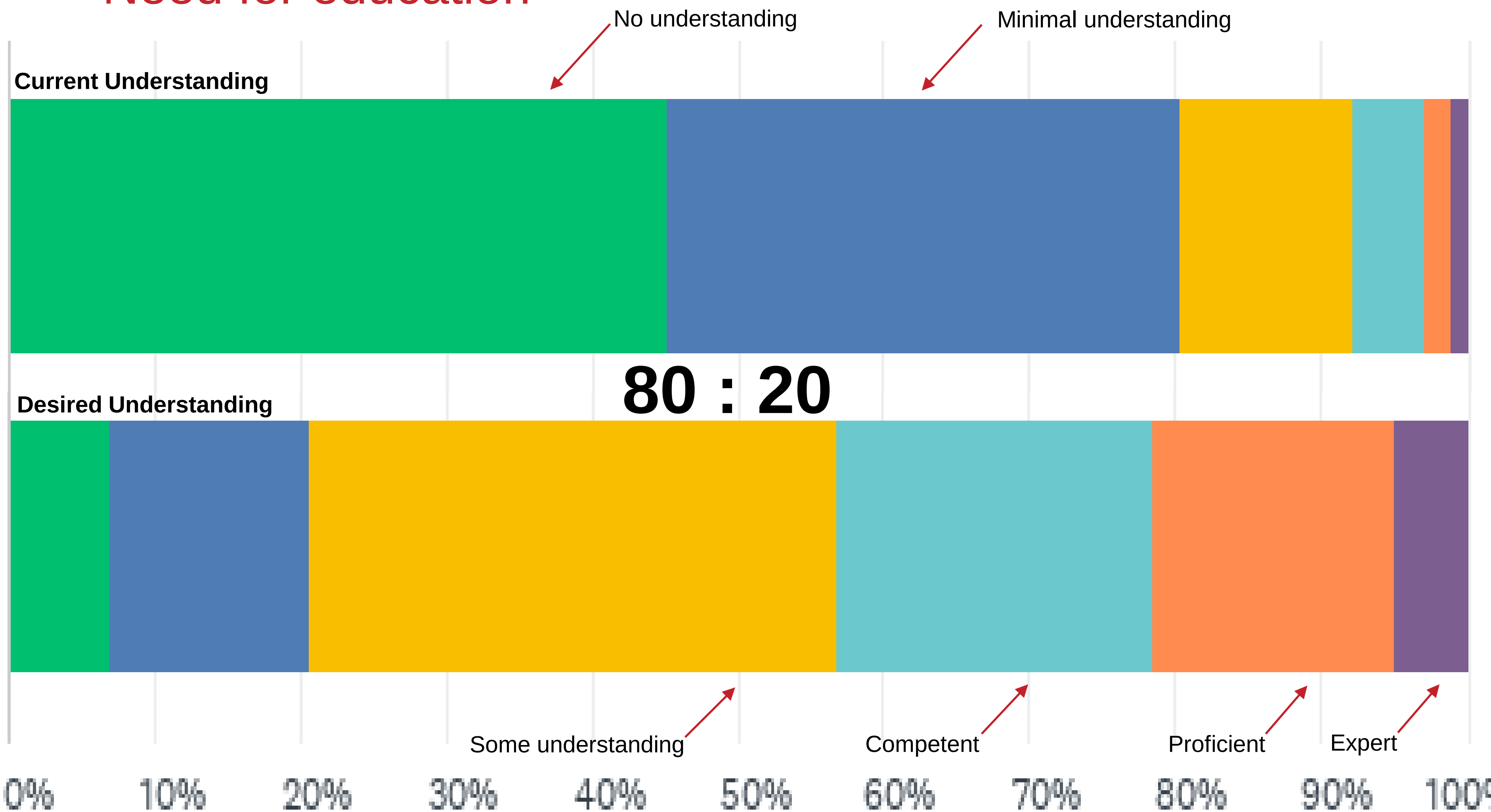
Recent Survey

- Concerns needing addressing:
  - Medico-legal issues
  - Ethics
  - Data privacy
  - Data diversity
- Need for education



Recent Survey

- Need for education



Not Fall Behind Radiographers!

Radiography 26 (2020) 93–95

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Guest editorial

Artificial Intelligence and the Radiographer/Radiological Technologist Profession: A joint statement of the International Society of Radiographers and Radiological Technologists and the European Federation of Radiographer Societies

BJR

Received: 30 September 2019 | Revised: 29 November 2019 | Accepted: 04 December 2019

https://doi.org/10.1259/bjr.20190840

Cite this article as:
Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. *Br J Radiol* 2020; **93**: 20190840.

REVIEW ARTICLE

Artificial intelligence in diagnostic imaging: impact on the radiography profession

[1]MARYANN HARDY, PhD, MSc, BSc(Hons), DCR(R) and [2]HUGH HARVEY, MBBS BSc(Hons) FRCR MD(Res)

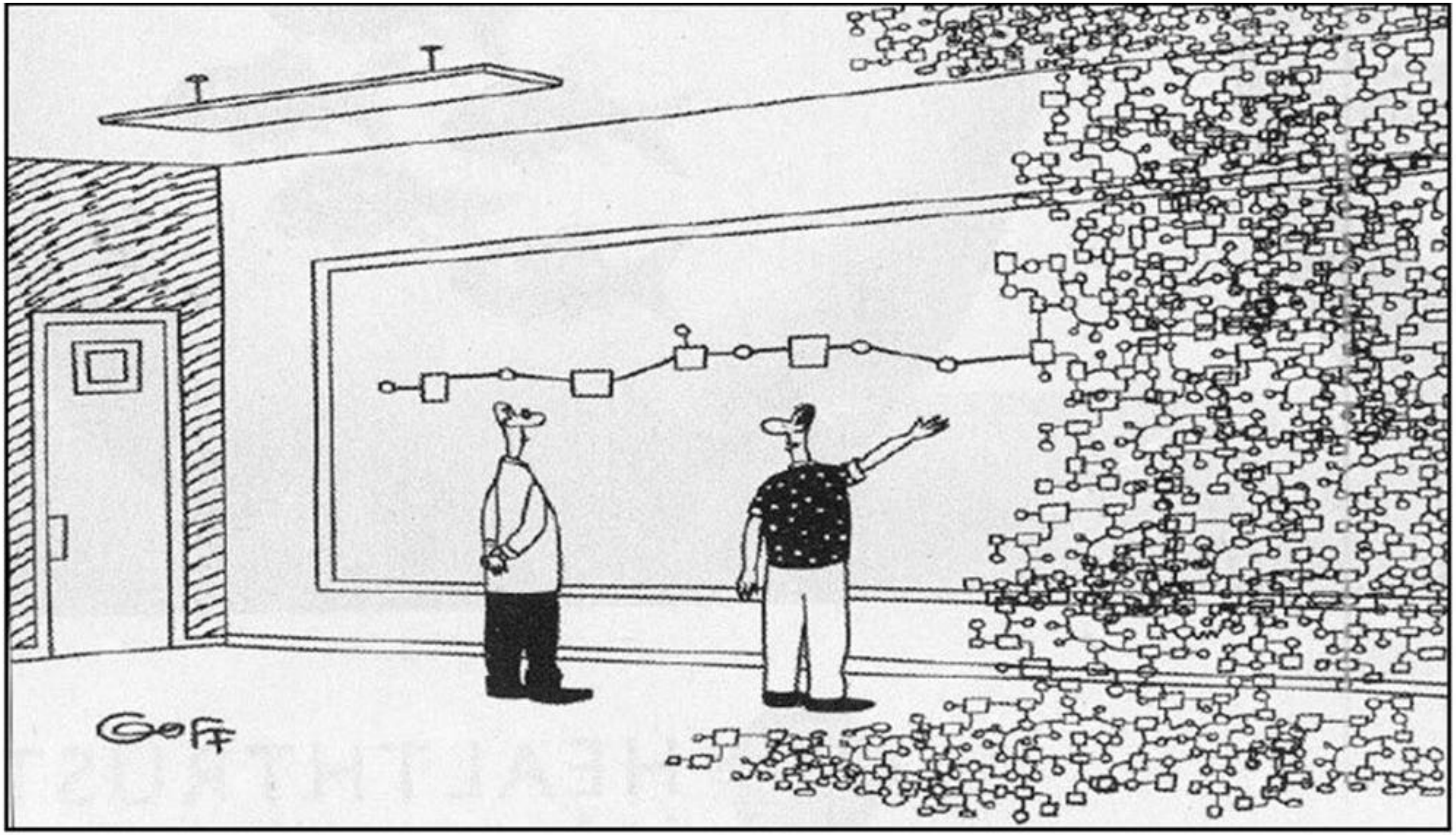
[1]University of Bradford, Bradford, England
[2]Hardian Health, Haywards Heath, UK

Address correspondence to: Professor Maryann Hardy
E-mail: *M.L.Hardy1@bradford.ac.uk*

Summary

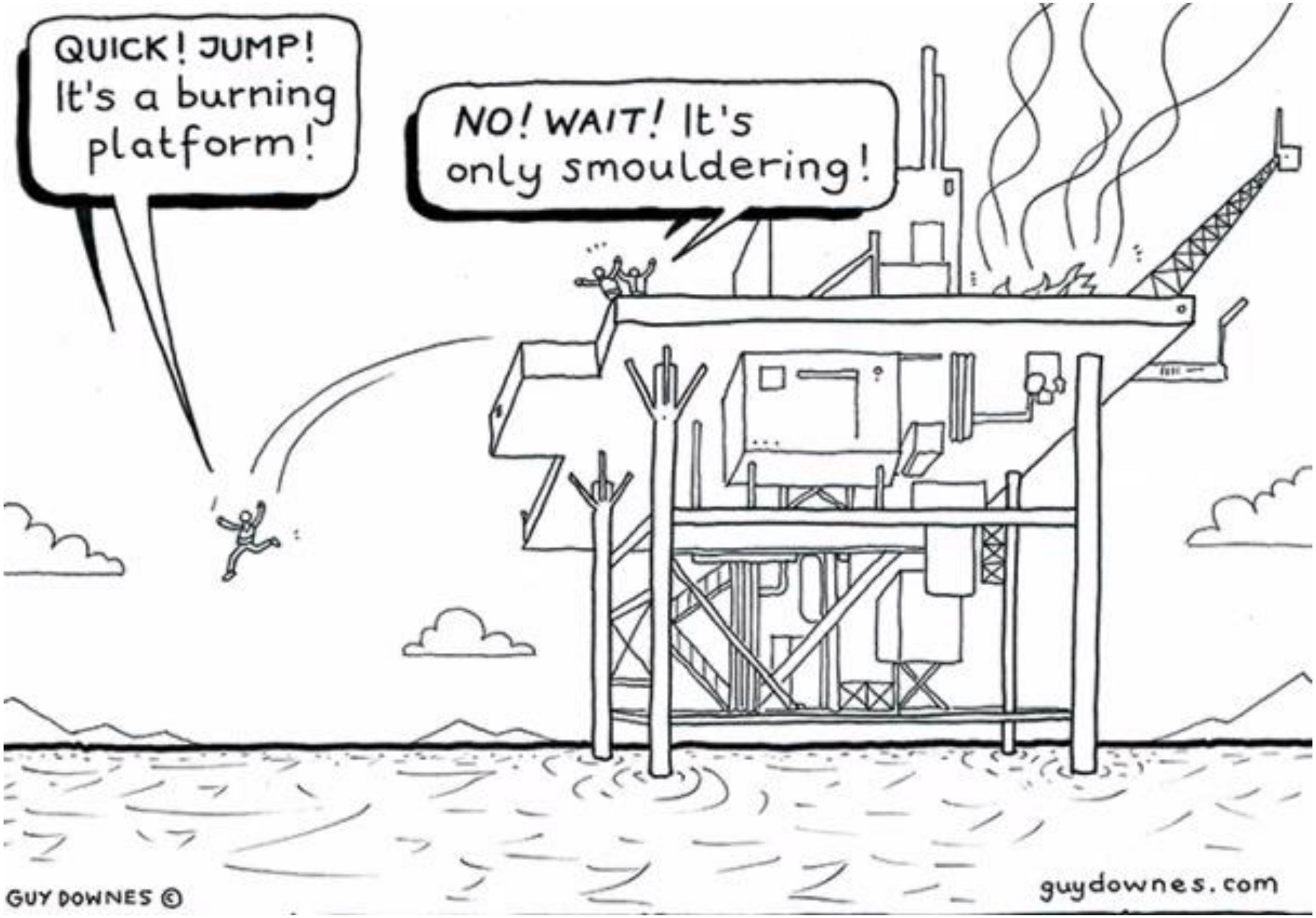
Reality of what NMTs need to know

NMT perception of learning AI



Thanks

Data is the new oil and AI the new electricity



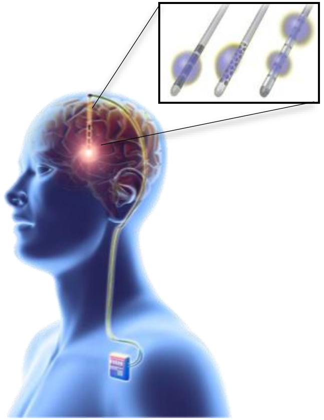
What FDA requires and desires for a nuclear medicine diagnostic AI Tool?

Dan Krainak, Ph.D.

Division of Radiological Health (DRH)
Office of Health Technologies ~~7~~ 8 (OHT8)
Office of Product Evaluation and Quality (OPEQ)
Center for Devices and Radiological Health (CDRH)
US Food and Drug Administration (FDA)

SNMMI Artificial Intelligence (AI) Summit
Virtually March 21-22, 2022

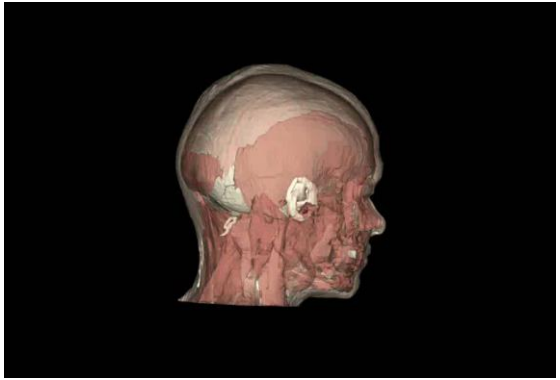
FDA’s Center for Devices and Radiological Health (CDRH)



.. 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…

…We provide consumers, patients, their caregivers, and providers with understandable and accessible science-based information about the products we oversee…

We facilitate medical device innovation by advancing regulatory science, providing industry with predictable, consistent, transparent, and efficient regulatory pathways, and assuring consumer confidence in devices marketed in the U.S.



Nuclear Medicine Devices

- Not diagnostic radiopharmaceuticals (*those are drugs*)
- Include imaging hardware and software
  - Acquisition hardware and software
  - Post-processing software
- Generally, a mixture of Class I and Class II
- Most post-processing software is Class II

Risk-based approach to device classification

Classification depends upon the degree of regulation necessary to provide reasonable assurance of safety and effectiveness

Class I: low risk, general controls
Class II: moderate risk, general controls + special controls
Class III: high risk, general controls + premarket approval

Guidance for Industry

Guidance for the Submission of Premarket Notifications for Emission Computed Tomography Devices and Accessories (SPECT and PET) and Nuclear Tomography Systems

Document issued on: December 3, 1998

Artificial Intelligence (AI) Tool

- What is an Artificial Intelligence (AI) tool in the context of nuclear medicine?
  - *If there's software, there might be AI*
- What are the regulatory expectations for AI Tools?
  - *"It depends"*

FDA on AI

FDA STATEMENT

Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices

April 2019

Good Machine Learning Practice for Medical Device Development: Guiding Principles

October 2021

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

April 2019

WORKSHOP

Public Workshop – Evolving Role of Artificial Intelligence in Radiological Imaging

February 25-26, 2020

Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices | FDA

List of more than 300 entries and growing

WORKSHOP

Virtual Public Workshop – Transparency of Artificial Intelligence/Machine Learning-enabled Medical Devices

October 14, 2021

Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan

January 2021

More to come … we're thinking about it

Radiological imaging – AI to date

- Computer assisted detection/diagnosis/triage
  - Task-specific
  - Identifies, marks, highlights, categorizes, characterizes, notifies, priorities, etc.
  - Intended to augment or improve physician performance
- Segmentation
  - Outline normal and/or abnormal features
- Acquisition optimization
  - For example, patient positioning, FOV optimization, hardware parameter selection
- Image enhancement
  - For example, image reconstruction, denoising
- Quantitative imaging
  - For example, ejection fraction based on ultrasound images

Classification of AI

- Generally, devices with AI follows the classification of the technology regulated without AI
- Again – most devices with AI in the nuclear medicine diagnostic space are anticipated to be Class II devices – require 510k notification

510(k) premarket notifications

Substantially equivalent (SE) (21 CFR 807.100(b)):

same intended use AND same technological characteristics

OR

same intended use AND different technological characteristics (e.g., change in material, design, energy source, software) AND these differences do not raise different questions of safety and effectiveness

“Tool type” claims

- For most 510(k) imaging devices, CDRH requests that sponsors provide validation consistent with the technological characteristics and intended use of the device
- Tool type intended use permit device manufacturers to make medical devices available to the community faster
- Tool claims encourage <u>clinical testing</u> of specific intended uses not called out in the indication for use statement by the <u>clinical community</u>

Quantitative tools and computer-aided ___

- Tool example
  - Calculate relative SUV (quantitative analysis)
- Diagnostic intended use
  - Lesion identification and classification (benign/malignant) = CAD/intended use
  - Disease status classification = CAD/intended use

Types of evidence to support substantial equivalence

- Phantoms (including both physical and digital reference objects) – *some challenges with AI*
- Simulations (realistic models)
- Clinical data
  - Reader studies
  - Validation of quantitative imaging

Transparency in AI & 510(k) Summaries

Validation datasets

- Summary test statistics or other test results including acceptance criteria or other information supporting the appropriateness of the characterized performance
- The number of individual patients images were collected from
- The number of samples, if different from above, and the relationship between the two
- Demographic distribution including
  - Gender
  - Age
  - Ethnicity
- Information about clinical subgroups and confounders present in the dataset
- Information about equipment and protocols used to collect images
- Information about how the reference standard was derived from the dataset (i.e. the “truthing” process)
- Description of how independence of test data from training data was ensured

Information about the training dataset should also be included as part of the device description

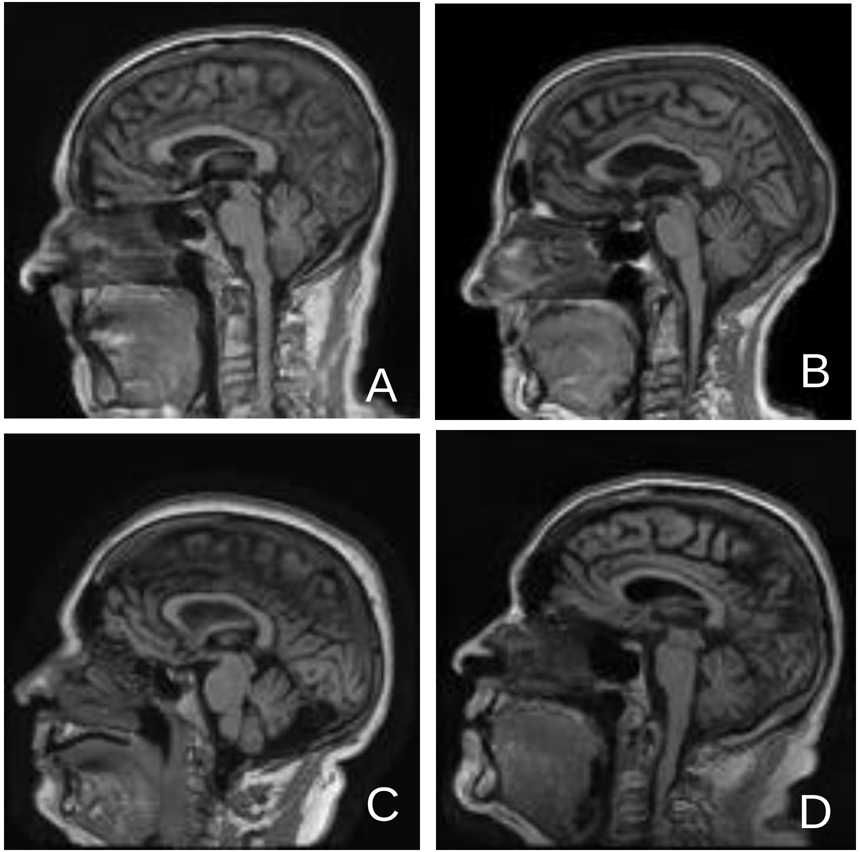
Image Quality

- In medical imaging, “image quality” is a measure of how much information an image gives us about a patient
  → Needed image quality is task-specific

  There are a variety of tracers, hardware, and software options that include user-selectable parameters for configuration, for many different target anatomies, for a variety of patient indications, and a variety of clinical tasks

- A clinical study to demonstrate diagnostic effectiveness is typically <u>not</u> requested as part of the premarket evaluation of PET hardware/software
  → What’s the right endpoint or the right clinical task?

Image Quality

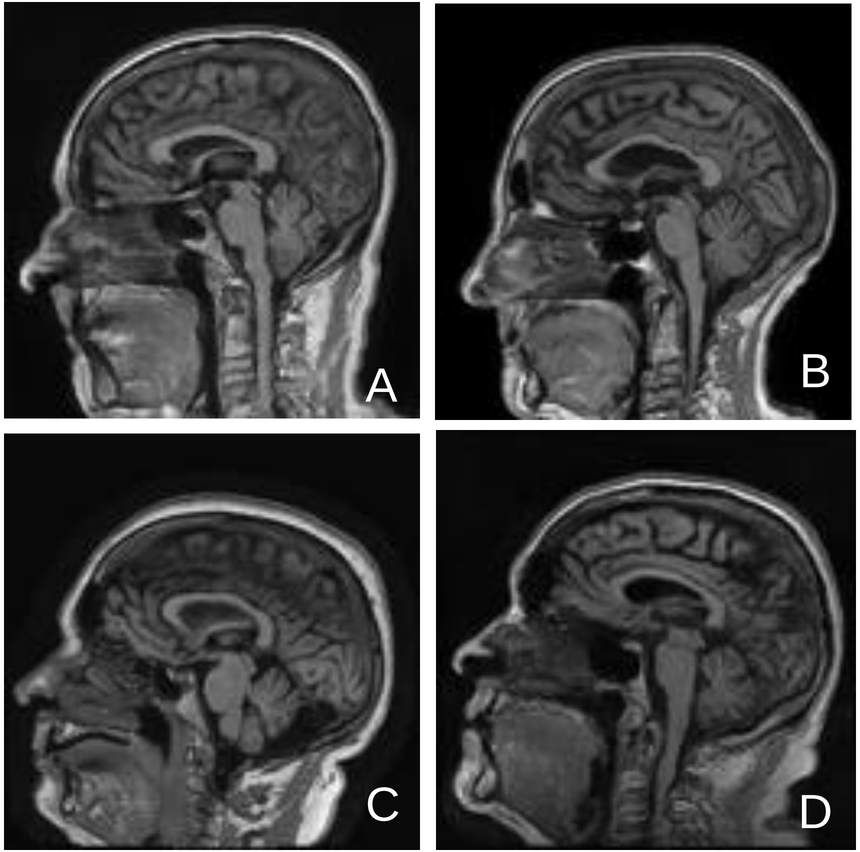


Are these images of adequate diagnostic quality?

Which of these images has the best quality?

NMR Images courtesy of Weimin Zhou and Mark Anastasio

Image Quality

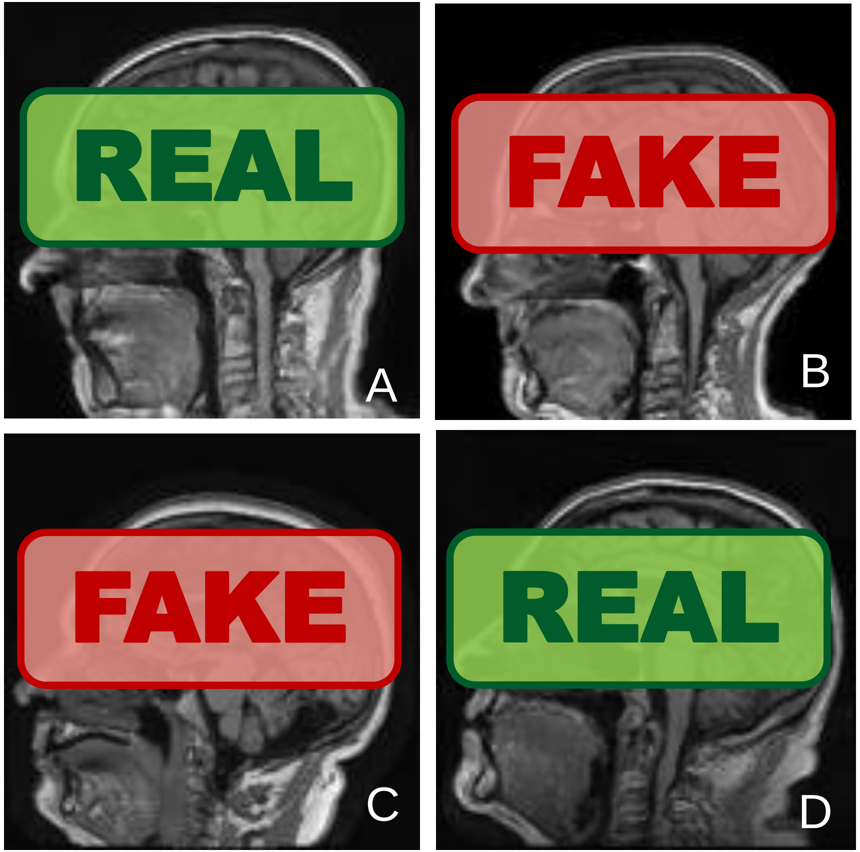


NMR Images courtesy of Weimin Zhou and Mark Anastasio

CAUTION: Image quality may be hard to identify

- Two of these real MR images give information about what is inside two real patients.
- Two of these images are generated by deep-learning neural networks, are completely fake, give no information about any patient, and therefore have NO image quality.

Image Quality



NMR Images courtesy of Weimin Zhou and Mark Anastasio

CAUTION: Image quality may be hard to identify

- Two of these real MR images give information about what is inside two real patients.
- Two of these images are generated by deep-learning neural networks, are completely fake, give no information about any patient, and therefore have NO image quality.

Nuclear Medicine AI

- Have not cleared any end-to-end black box image reconstruction methods
- Software as a medical device (SaMD) and software in a medical device (SiMD)
- Some features cleared include
  – Denoising, post-processing filters (image space)
  – Methods integrated into iterative reconstruction (sinogram)

Summary

- AI has many different meanings in the context of nuclear medicine
- Most hardware and some software are based on “tool type” claims – but some require more rigorous evaluation
- FDA emphasizes transparency in the context of AI based on feedback from the community

GUIDANCE DOCUMENT

Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program

Guidance for Industry and Food and Drug Administration Staff

JANUARY 2021

[Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program | FDA]

FDA U.S. FOOD & DRUG ADMINISTRATION

NCI Imaging Data Commons

Keyvan Farahani, PhD
Center for Biomedical Informatics and Information Technology

farahani@nih.gov
datascience.cancer.gov

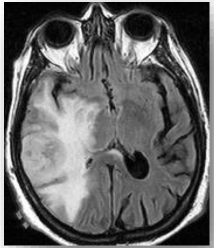
NIH NATIONAL CANCER INSTITUTE

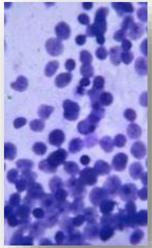
SNMMI AI Summit 2022

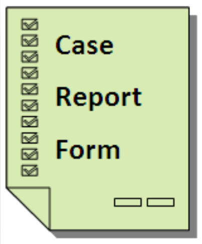
Outline

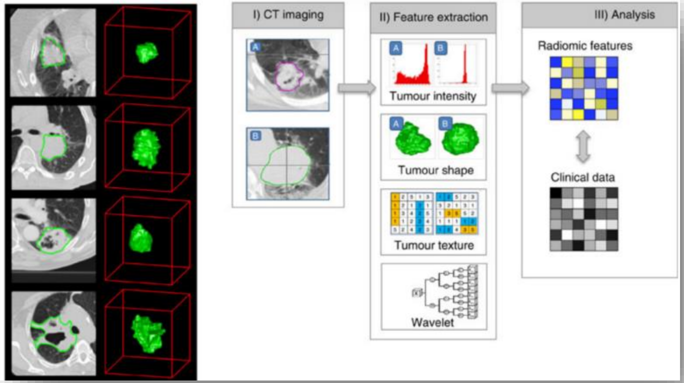
- Cancer Research Data Commons (CRDC)
- Imaging Data Commons (IDC)
- IDC imaging use cases

TCIA Overview









http://cancerimagingarchive.net

- New collection proposals are reviewed by the TCIA Advisory Group for quality and utility
- 140+ collections - data from > 55,000 subjects available for download
  - Preclinical imaging for multiple species
- Covers radiology, radiation therapy, and pathology image modalities
- Wide variety of cancers + phantoms
- Most have associated supporting data
  - Demographics/outcomes/therapy
  - Image Analyses (annotations, segmentations, features)
  - Links to Genomics/Proteomics
- REST API
- TCIA publishes data (DOI’s link to collections) and is a recognized repository for a growing number of scientific journals.

National Cancer Data Ecosystem for Sharing and Analysis

Cancer Moonshot℠

Overarching goals – Jan. 2016

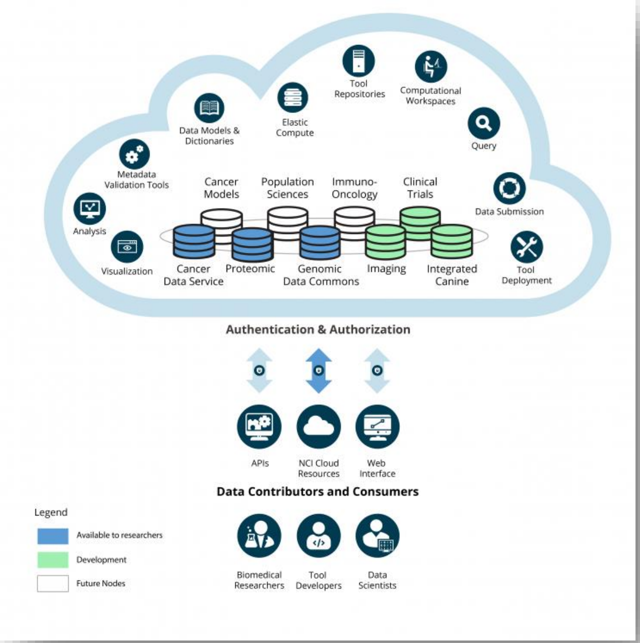
- **Accelerate progress in cancer, including prevention & screening**
  - From cutting edge basic research to wider uptake of standard of care
- **Encourage greater cooperation and collaboration**
  - Within and between academia, government, and private sector
- **Enhance data sharing**

- **Build a National Cancer Data Ecosystem**
  - Essential underlying data science infrastructure, standards, methods, and portals for the Cancer Data Ecosystem
  - Enhanced cloud-computing platforms
  - Services that link disparate information, including clinical, image, and molecular data
  - Establish sustainable data governance to ensure long-term health of the Ecosystem.
  - Develop standards and tools so that data are interoperable.

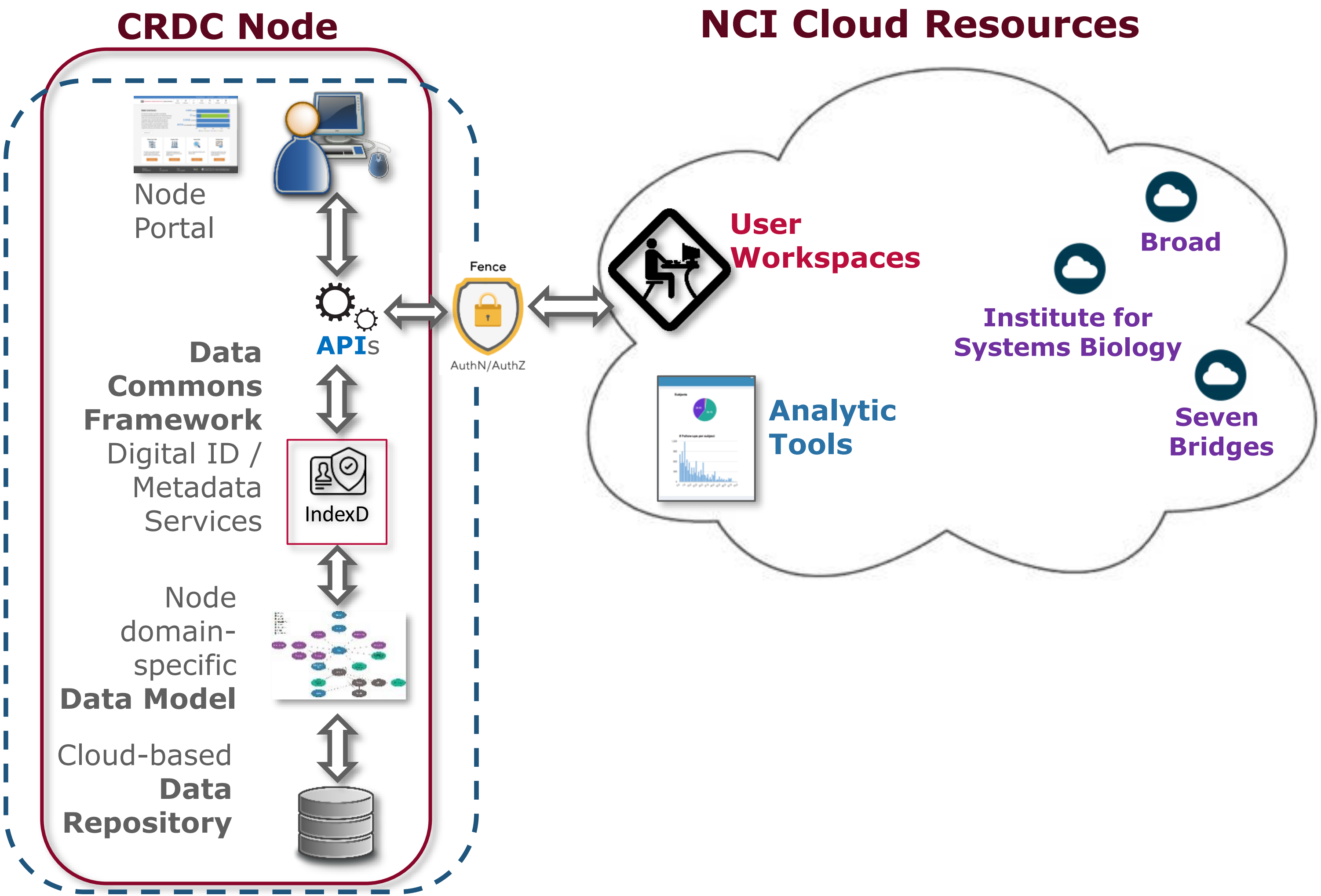
Cancer Research Data Commons (CRDC)

A data science infrastructure to connect repositories, analytical tools, and knowledge bases

- **Virtual, expandable, secure research infrastructure**
- **Storage and elastic compute**
- **Analysis, sharing, and archival of results**
- **Cross-domain analysis of large datasets**



datacommons.cancer.gov



CRDC Node
NCI Cloud Resources
Node
Portal
Fence
APIs
AuthN/AuthZ
Data
Commons
Framework
Digital ID /
Metadata
Services
IndexD
Node
domain-
specific
Data Model
Cloud-based
Data
Repository
User
Workspaces
Analytic
Tools
Broad
Institute for
Systems Biology
Seven
Bridges

The NCI Cloud Resources

Three resources connecting NCI data and compute in the cloud

- Access to large cancer data sets without need to download
- Access to workspaces, analysis tools, and pipelines
- Ability for researchers to bring their own data and tools

Data

- Access and analyze data from a dozen genomics, proteomics, and imaging datasets without downloading
- Upload your data to the cloud

Compute

- Perform large scale analysis using the elastic compute of commercial cloud platforms
- Upload your tools to the cloud, create your own workflows

Security

- dbGaP-authorized users can connect to controlled access datasets
- Systems meet strict Federal security guidelines

NCI Cloud Resources

Institute for Systems Biology

isb-cgc.org

FireCloud POWERED BY Terra

firecloud.terra.bio

cancergenomicscloud.org

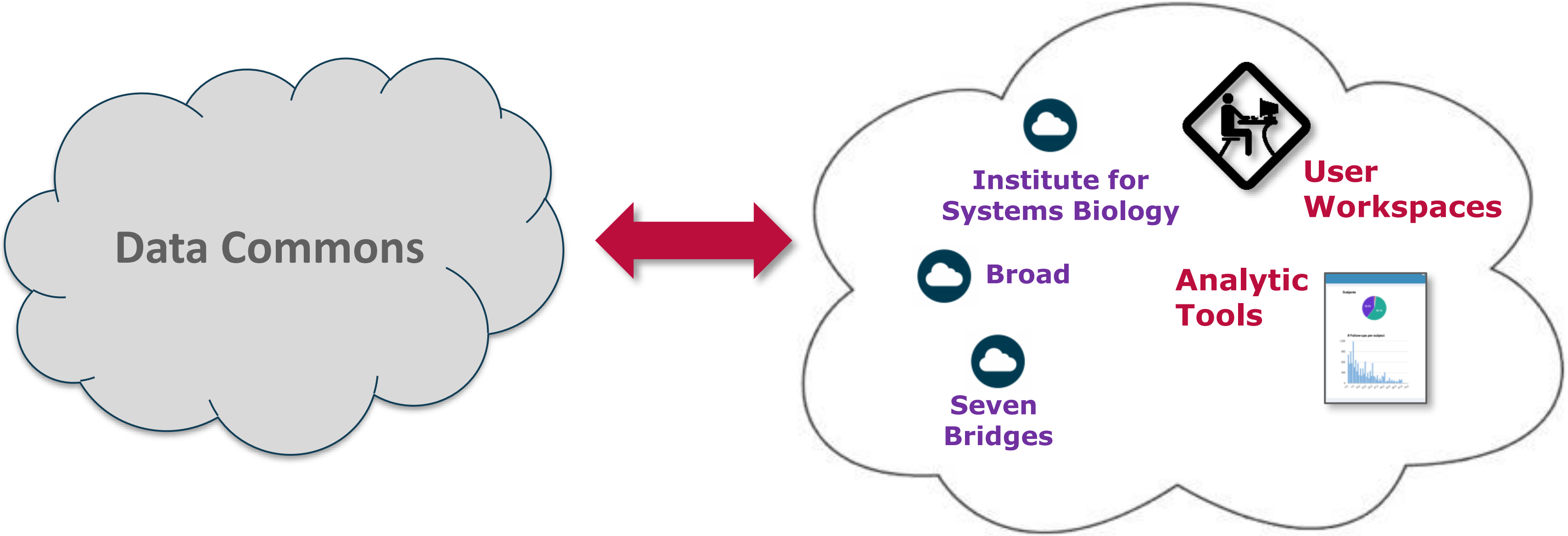
aws

#NCICloud

Cancer Research Data Commons

Cloud Repositories

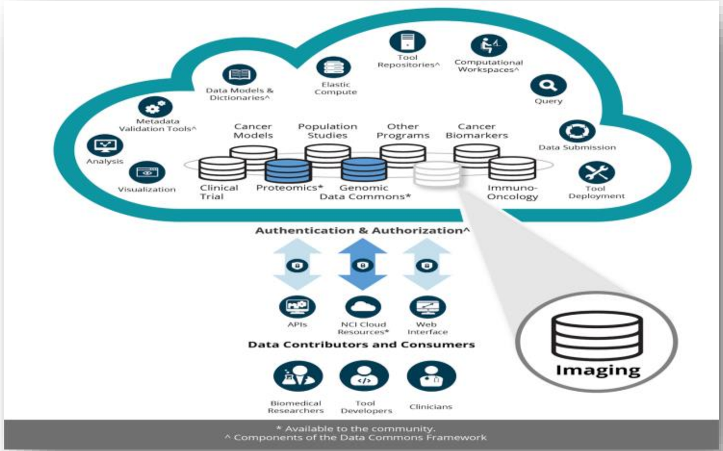
Cloud Resources



NCI Imaging Data Commons (IDC)

Cloud resource that connects researchers with:

- *Cancer image collections*
- *Robust infrastructure with imaging data, metadata and experimental metadata from disparate sources*
- *Resources for searching, identifying and viewing images*
- *Additional data types in other CRDC nodes*
- *Connectivity to NCI Cloud Resources for imaging and multi-modal cloud computations*



Implementation:
- Google Cloud Platform
- OHIF viewer
- Non-restrictive Open Source
- DICOM as prime standard

Production release: September 2021

IDC leadership

Ron Kikinis

Andrey Fedorov

Hugo Aerts

Mass General Brigham

Bill Longabaugh

ISB

David Pot

GDIT

David Clunie

PixelMed

Steve Pieper

Isomics, Inc.

Rob Lewis

RADICAL IMAGING

André Homeyer

Fraunhofer MEVIS

Todd Pihl

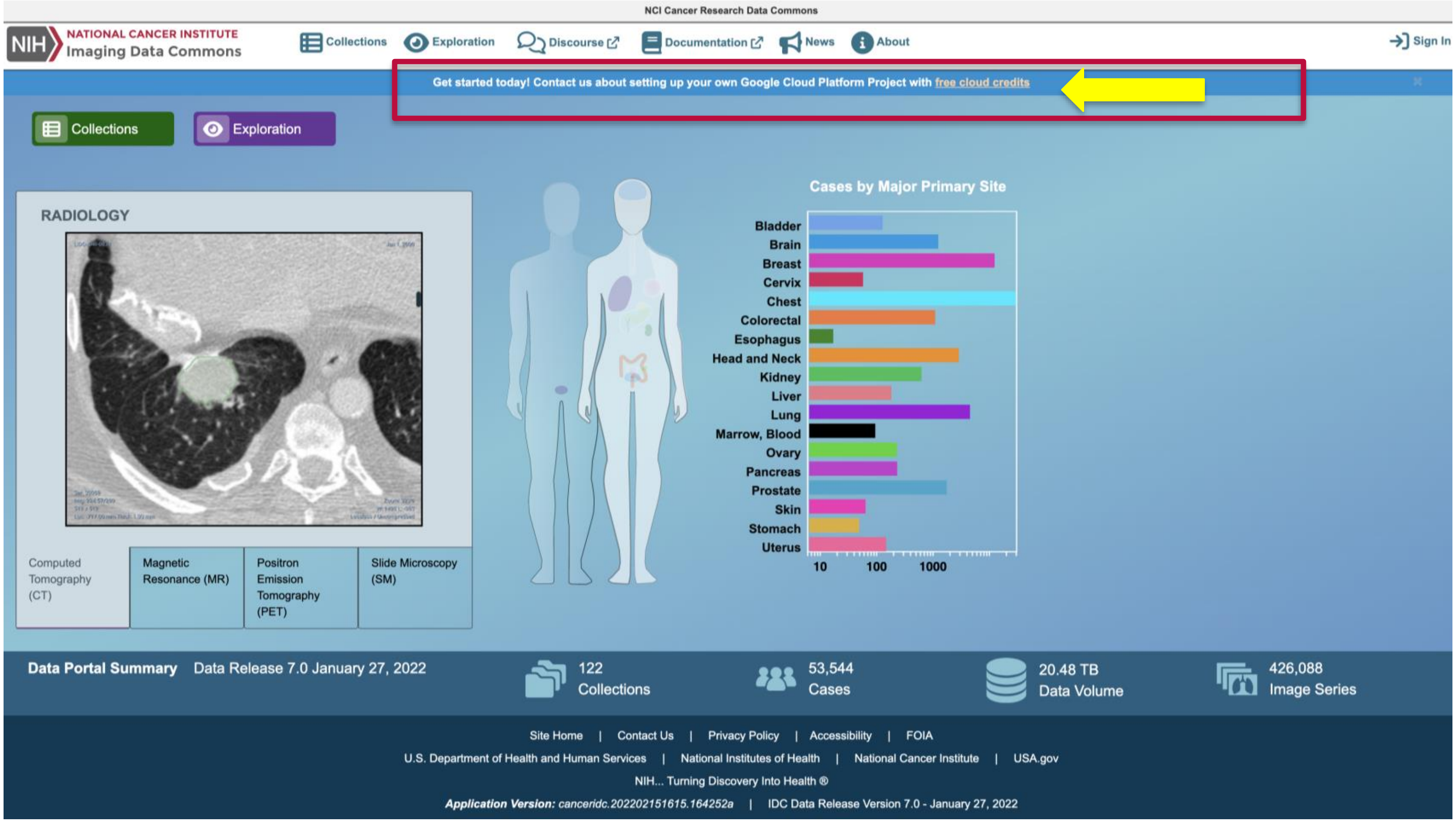
Ulrike Wagner

Frederick National Laboratory for Cancer Research

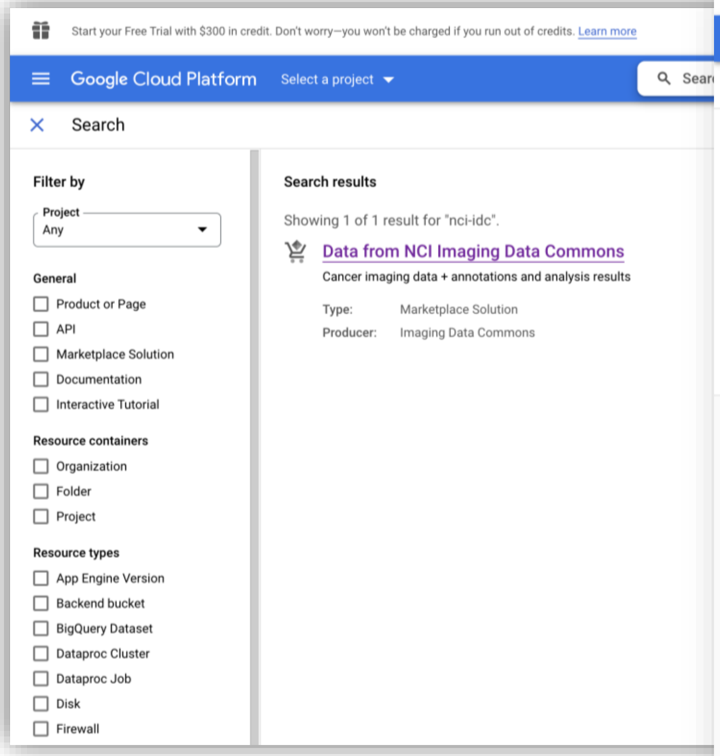
Keyvan Farahani

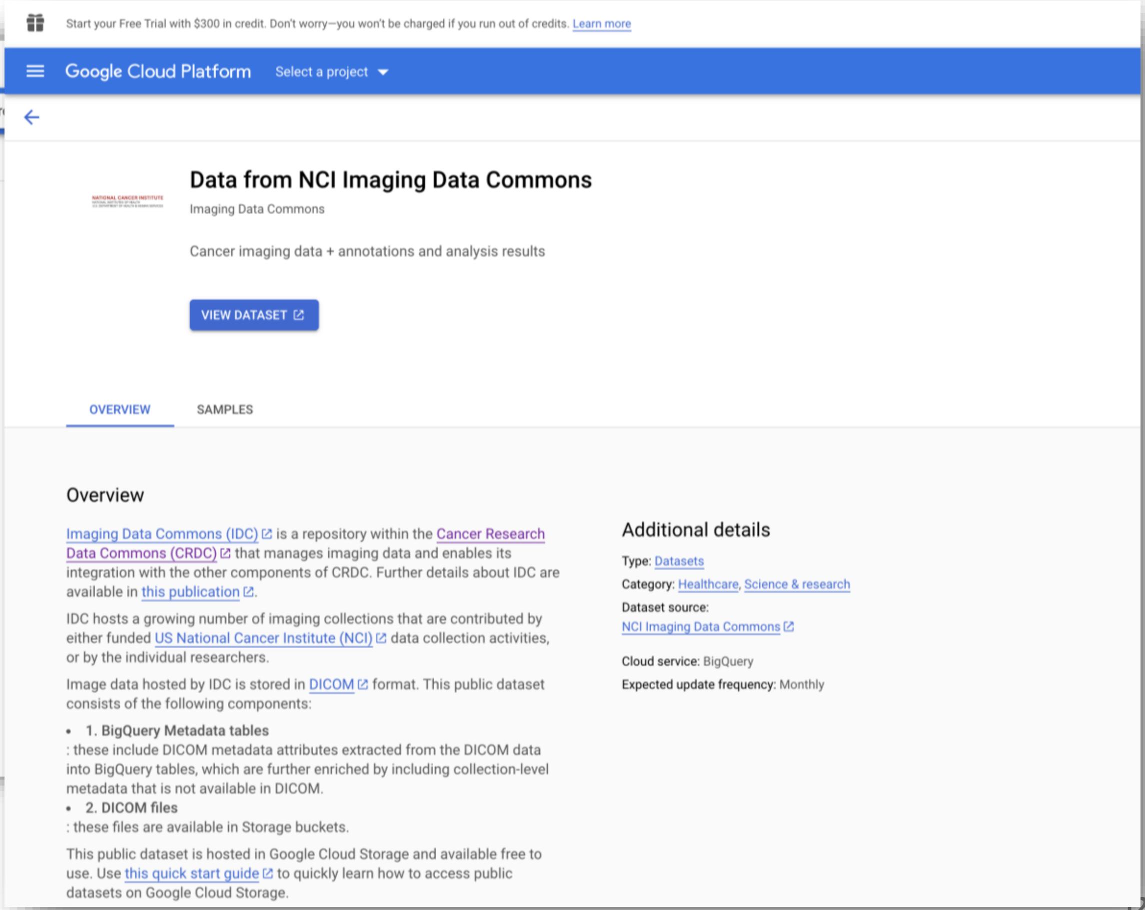
NIH NATIONAL CANCER INSTITUTE

IDC portal: imaging.datacommons.cancer.gov



IDC in Google Public Dataset Program





https://console.cloud.google.com/

nci-idc

Medical Image De-Identification Initiative (MIDI)*
Projects 1 and 2

Overall Goal: To address the need for a scalable, automated, AI-based image de-ID

Data w/ synthetic PHI
[MIDI Project 1]

De-ID pipeline and performance evaluation
[MIDI Project 2]

Rutherford, et. al., Nature Sci Data 2021

*MIDI is independent of IDC.

Artificial Intelligence & IDC

- IDC can play a central role by providing data to enable end-to-end transparent and reproducible AI pipelines for cancer imaging.
- Easy access to high quality, standardized, de-identified imaging and metadata in IDC that can be combined with fully reproducible AI pipelines in cloud based environments.
- AI researchers are empowered to reproduce published results, provide materials for research, training and education purposes, as well as guide overall developments of the IDC platform.
- Selected AI use cases for several clinical scenarios in cancer imaging are being developed by IDC and collaborators to highlight these capabilities.

IDC AI workflow

Data Retrieval → Data Investigation → Environment Set-up → Experiment and Results Analysis

Google Cloud Storage and BigQuery

Google Cloud Platform and IDC Viewer

Google VMs, AI Notebooks and Google Colab

Use Case I: Lung Cancer Prognosis

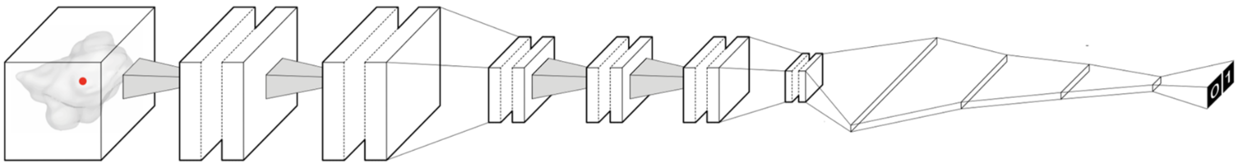
AI to predict Non-small Cell Lung Cancer Patient 2-year survival

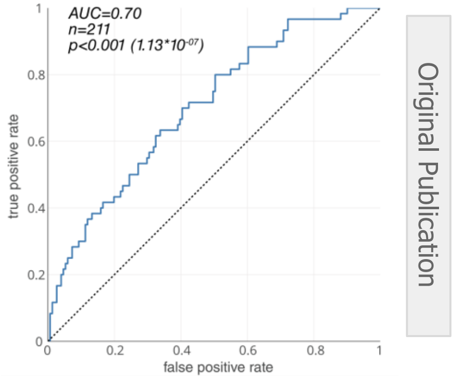
- Trained and tuned on institutional data
- Tested on public data (NSCLC-Radiomics) available on IDC

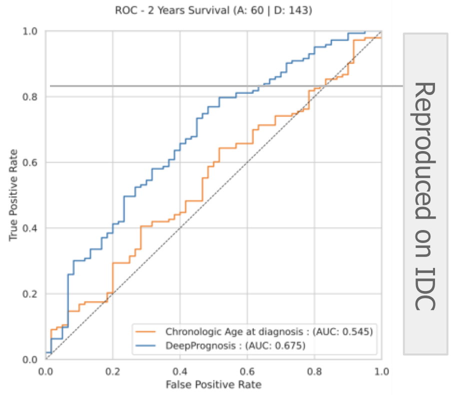
Challenges

- Replicate exactly the pre-processing pipeline
- Dataset evolved in the meantime (new segmentations)

➔ Successfully replicated the results in Hosny et al.







ImagingDataCommons/IDC-Examples/notebooks/nsclc-radiomics

Slide courtesy of Hosny, Fedorov, Aerts (Mass General Brigham)

Hosny et Al. - *Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study (PLOS Medicine, 2018)*

Use Case II: Thoracic OAR Segmentation

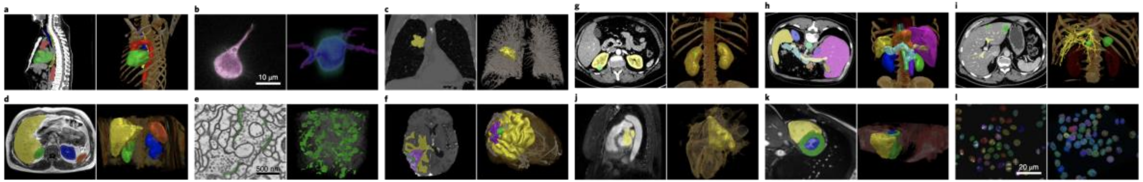
Used nnU-Net, a collection of AI models for biomedical image segmentation, to segment previously unseen IDC data

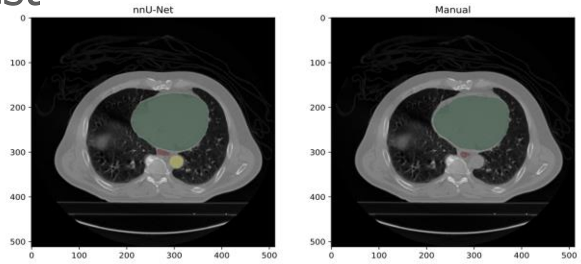
- Thoracic organs at risk (OAR) segmentation from chest CT for radiotherapy planning

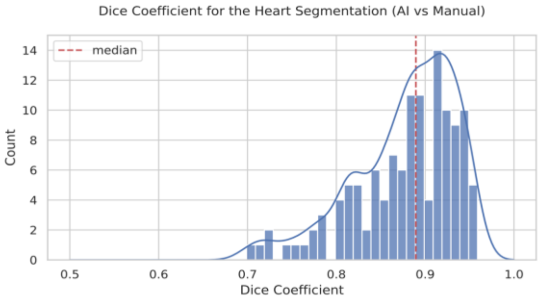
Challenges:

- Set up the pipeline correctly, integration with IDC
- Pre- and post-process the data

➔ Successfully integrated different nnU-Net models with the IDC data on Google Cloud Platform







ImagingDataCommons/IDC-Examples/notebooks/thoracic_oar_demo.ipynb

Slide courtesy of Bontempi, Fedorov, Aerts (Mass General Brigham)

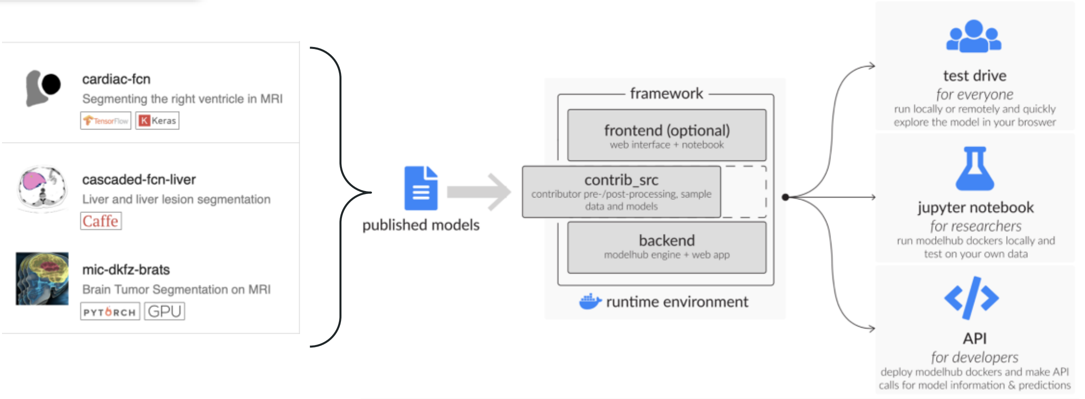
Isensee et Al. - *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation (Nature Methods, 2021)*

Artificial Intelligence & IDC - What's Next

Continue to investigate how to promote transparency, reproducibility and reusability

➔ Promoting the usage of tools that allow easy deployment and testing of AI pipelines for biomedical image analysis

ModelHub.ai



monai.io

MONAI
MEDICAL OPEN NETWORK FOR AI

dockstore.org

Dockstore
Create, Share, Use

IDC Use Cases

- Essential to promote utilization of IDC/CRDC infrastructure and standards toward:
  - Development of novel AI/ML tools
    - Various applications in imaging – detection, diagnosis, and treatment planning/monitoring
  - Training of next generation of imaging data scientists
- Additional cloud-credits may be available to support novel developments

farahani@nih.gov

datascience.cancer.gov

www.cancer.gov

www.cancer.gov/espanol

Big Issues in Big Data Facing NCI

Workforce and career development

EHR Mining

Storage – What? How Long? Cloud?

Security, privacy and de-identification

Use of challenges / prizes

From Dr. Sharpless (NCI Director), May 2018

Our Experience with Challenges - a brief history

2013 – Prostate MRI segmentation [IEEE-ISBI]

201 ... E

201 2 ... tation (CPM

2015 –

2015 –

2016 –

2017 – ... on Kaggle]

2017 – ... SPIE-AAPM-F

2018 –

2018 –

2018 –

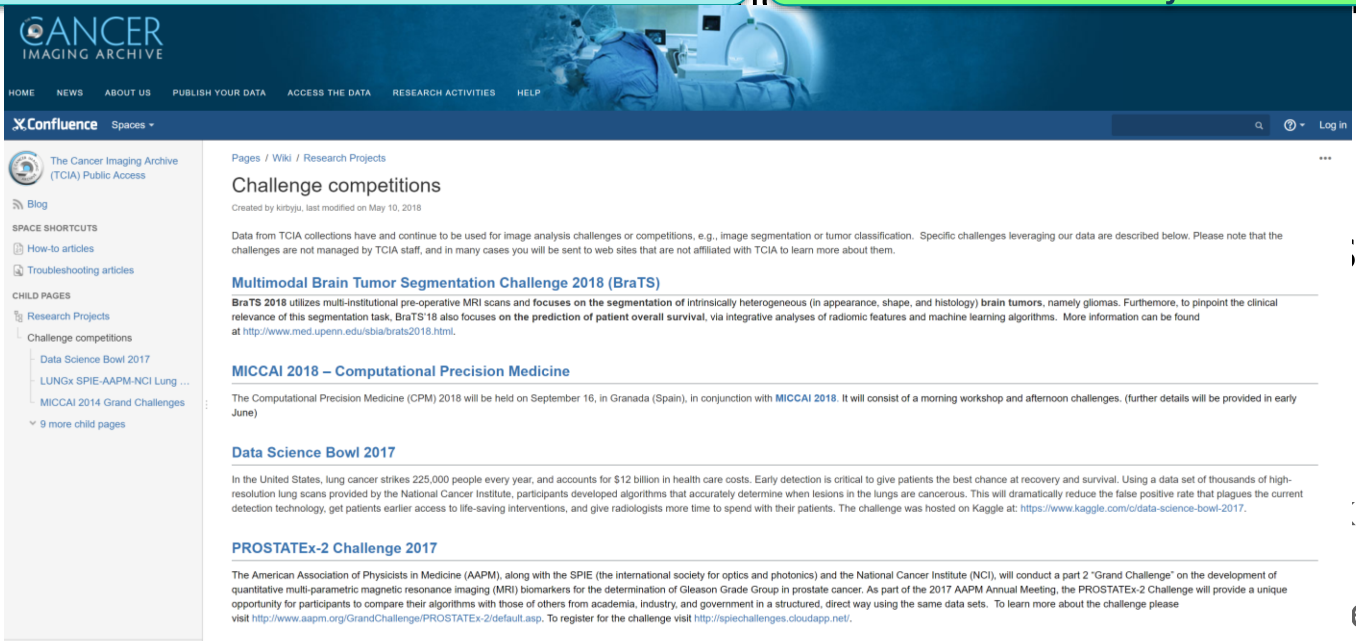
2019 – ... s]

2019 –

2019 – ... er, RSNA]

Computational Precision Medicine

QIN Challenges & Collaborative Projects


CANCER IMAGING ARCHIVE

HOME NEWS ABOUT US PUBLISH YOUR DATA ACCESS THE DATA RESEARCH ACTIVITIES HELP

Confluence Spaces Log in

The Cancer Imaging Archive (TCIA) Public Access

Blog

SPACE SHORTCUTS

How-to articles

Troubleshooting articles

CHILD PAGES

Research Projects

Challenge competitions

Data Science Bowl 2017

LUNGx SPIE-AAPM-NCI Lung ...

MICCAI 2014 Grand Challenges

9 more child pages

Pages / Wiki / Research Projects

Challenge competitions

Created by kirbyju, last modified on May 10, 2018

Data from TCIA collections have and continue to be used for image analysis challenges or competitions, e.g., image segmentation or tumor classification. Specific challenges leveraging our data are described below. Please note that the challenges are not managed by TCIA staff, and in many cases you will be sent to web sites that are not affiliated with TCIA to learn more about them.

Multimodal Brain Tumor Segmentation Challenge 2018 (BraTS)

BraTS 2018 utilizes multi-institutional pre-operative MRI scans and **focuses on the segmentation of** intrinsically heterogeneous (in appearance, shape, and histology) **brain tumors**, namely gliomas. Furthemore, to pinpoint the clinical relevance of this segmentation task, BraTS'18 also focuses **on the prediction of patient overall survival**, via integrative analyses of radiomic features and machine learning algorithms. More information can be found at http://www.med.upenn.edu/sbia/brats2018.html.

MICCAI 2018 – Computational Precision Medicine

The Computational Precision Medicine (CPM) 2018 will be held on September 16, in Granada (Spain), in conjunction with MICCAI 2018. It will consist of a morning workshop and afternoon challenges. (further details will be provided in early June)

Data Science Bowl 2017

In the United States, lung cancer strikes 225,000 people every year, and accounts for $12 billion in health care costs. Early detection is critical to give patients the best chance at recovery and survival. Using a data set of thousands of high-resolution lung scans provided by the National Cancer Institute, participants developed algorithms that accurately determine when lesions in the lungs are cancerous. This will dramatically reduce the false positive rate that plagues the current detection technology, get patients earlier access to life-saving interventions, and give radiologists more time to spend with their patients. The challenge was hosted on Kaggle at: https://www.kaggle.com/c/data-science-bowl-2017.

PROSTATEx-2 Challenge 2017

The American Association of Physicists in Medicine (AAPM), along with the SPIE (the international society for optics and photonics) and the National Cancer Institute (NCI), will conduct a part 2 "Grand Challenge" on the development of quantitative multi-parametric magnetic resonance imaging (MRI) biomarkers for the determination of Gleason Grade Group in prostate cancer. As part of the 2017 AAPM Annual Meeting, the PROSTATEx-2 Challenge will provide a unique opportunity for participants to compare their algorithms with those of others from academia, industry, and government in a structured, direct way using the same data sets. To learn more about the challenge please visit http://www.aapm.org/GrandChallenge/PROSTATEx-2/default.asp. To register for the challenge visit http://spiechallenges.cloudapp.net/.

Past collaborators in academia and industry

MICCAI

kaggle Competitions

Memorial Sloan Kettering Cancer Center

SageBionetworks

Booz | Allen | Hamilton

NVIDIA

FDA

Google Cloud

Stony Brook University

IN·DEI·NOMINE·FELICITER

Radboud University Nijmegen

UCL

AAPM American Association of Physicists in Medicine

MGH 1811

ljaf laura and john arnold foundation

SPIE.

IEEE EMB

IEEE Signal Processing Society

THE UNIVERSITY OF TEXAS MD Anderson ~~Cancer~~ Center

THE UNIVERSITY OF CHICAGO

Contact:

farahani@nih.gov

Paul Gruenberg

Patient with Metastatic Castrate Resistant Prostate Cancer

Recently traveled to Germany for Lu177-PSMA-617 therapy

Strategic and Financial Advisor to BMAF – A Grand Rapids-based Theranostics Center

Medical Imaging and Implementation Science in Dynamic Systems: An NIH Perspective

David Chambers, DPhil

Deputy Director for Implementation Science,

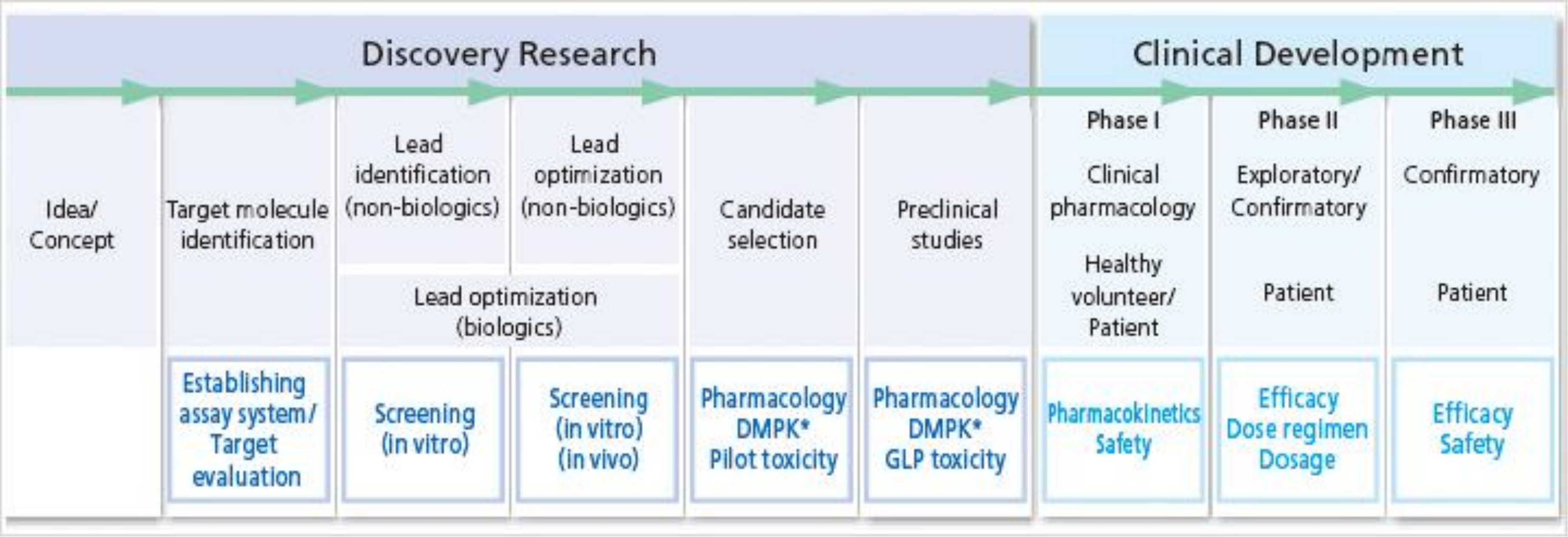
Division of Cancer Control & Population Sciences (DCCPS)

NIH NATIONAL CANCER INSTITUTE

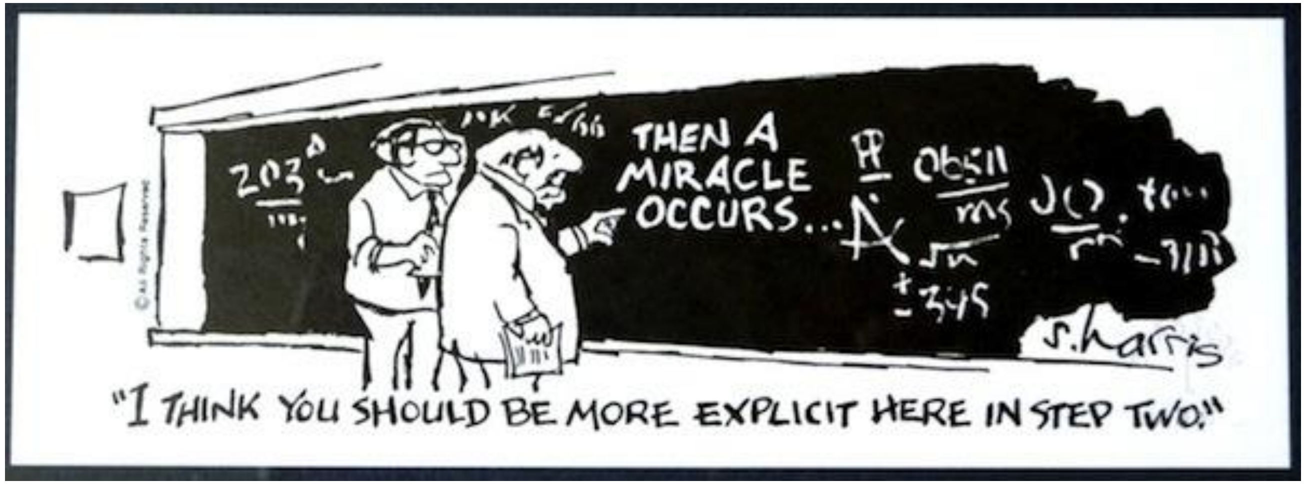
Session Outline

- What is Implementation Science and How Does it Relate to Medical Imaging?
- Key IS Activities/Resources
- Areas for Further Development

From Discovery to Delivery

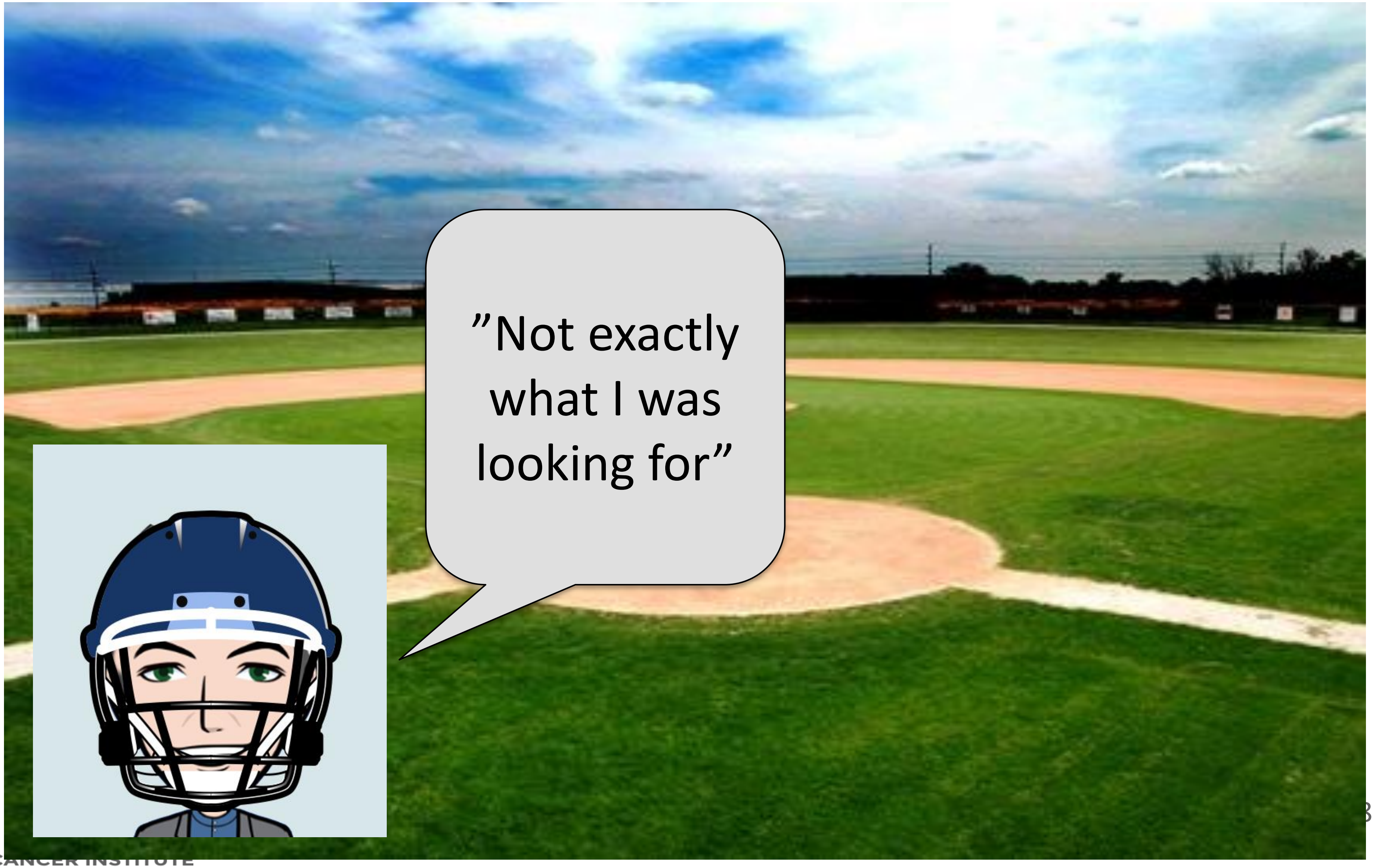


*Drug metabolism and pharmacokinetics



IF YOU BUILD IT, THEY MAY NOT COME

IF YOU BUILD IT, THEY MAY NOT COME



An AI-Driven Medical Imaging Intervention

- Is only so good as how and whether. . .
  - It is adopted?
  - Providers are trained to deliver it?
  - Trained providers choose to deliver it?
  - Eligible people receive?

If we assume 50% threshold for each step. . .

(even w/perfect access/adherence/dosage/maintenance)

Impact: .5*.5*.5*.5=6% benefit

Adapted from Glasgow, RE-AIM

More than Efficacy/Effectiveness

Glasgow, Vogt, & Boles (1999)

Key Terms

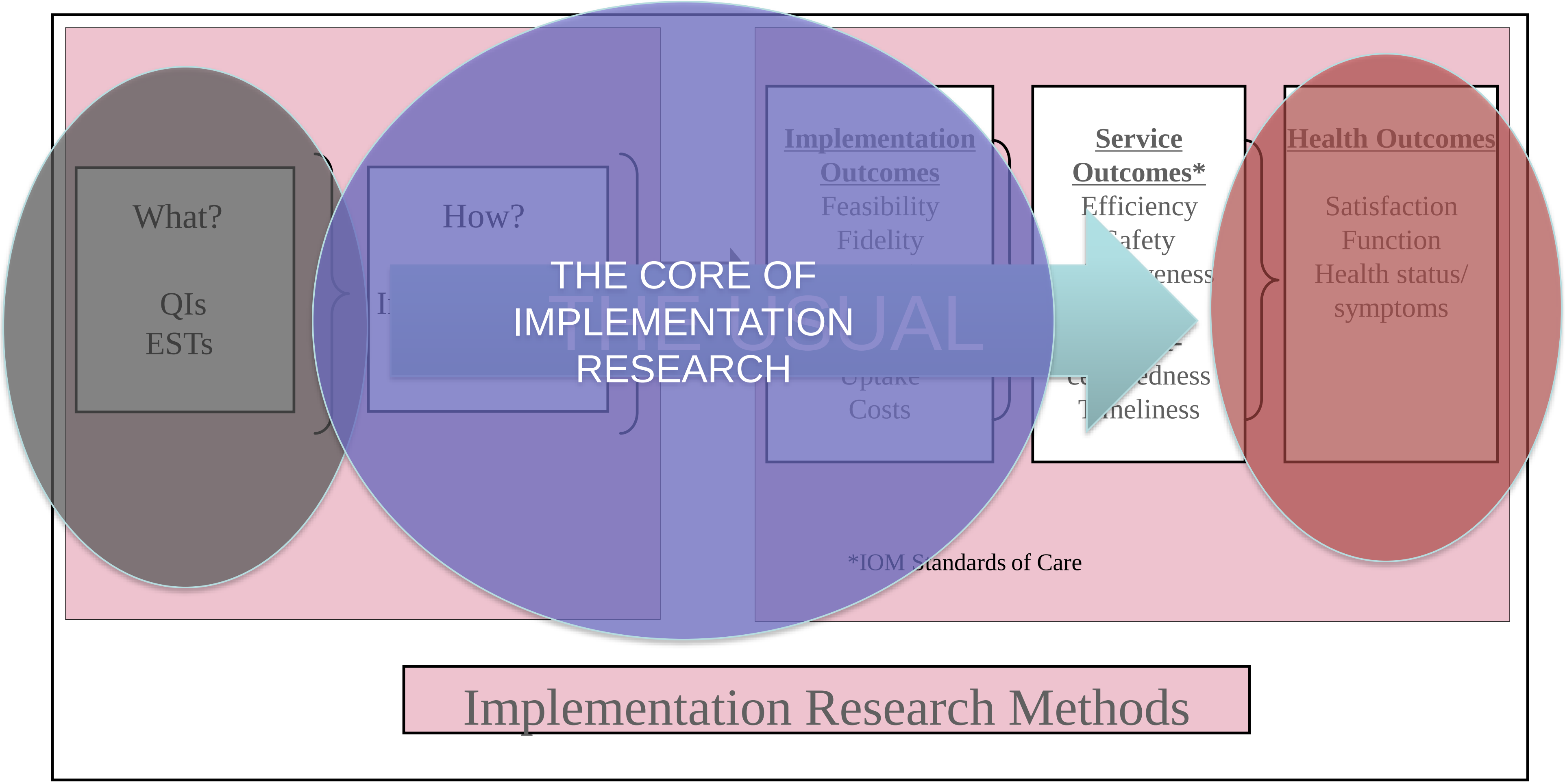
- ***Implementation Science*** is the study of methods to promote the integration of research findings and evidence into healthcare policy and practice.
- ***Dissemination research*** is the scientific study of targeted distribution of information and intervention materials to a specific public health or clinical practice audience. The intent is to understand how best to spread and sustain knowledge and the associated evidence-based interventions.
- ***Implementation research*** is the scientific study of the use of strategies to adopt and integrate evidence-based health interventions into clinical and community settings in order to improve patient outcomes and benefit population health.

Dissemination Research

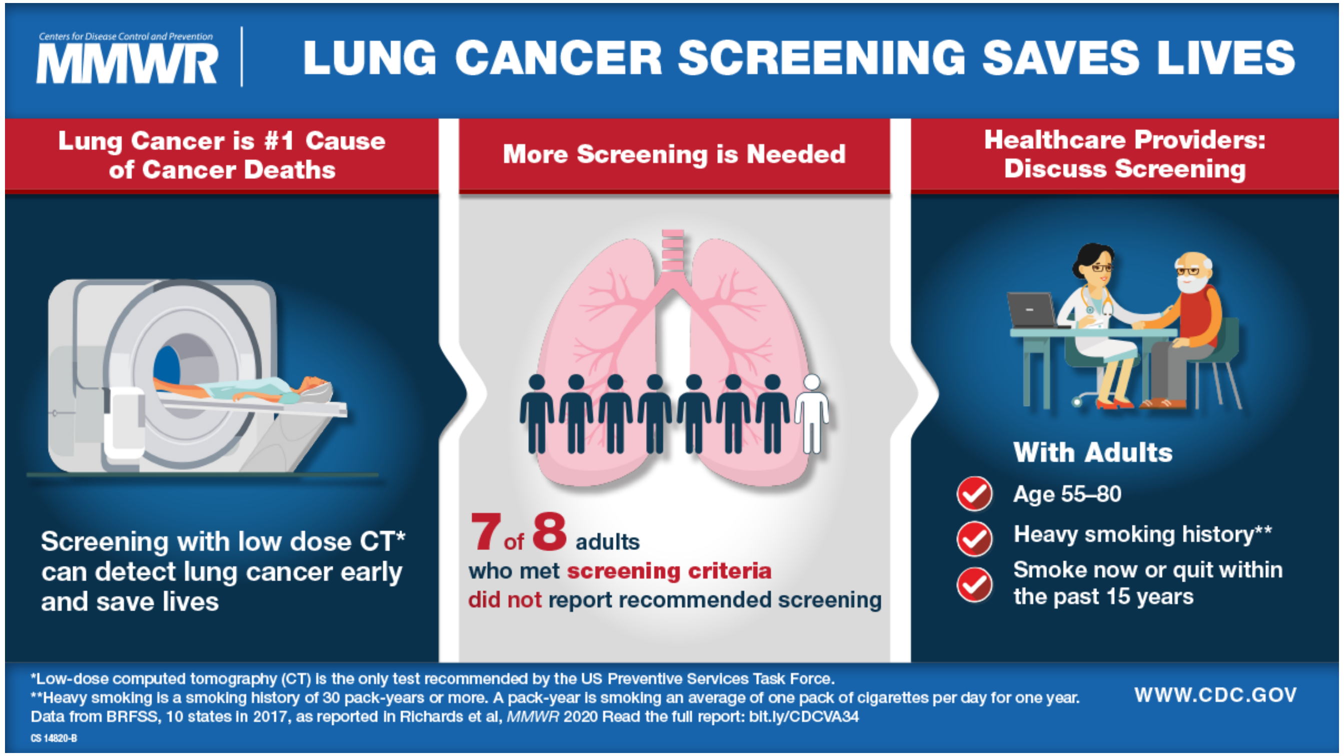
- How the “evidence” is created?
- Packaging
- Transmitting
- Receiving
- Turning Information into Action

- Many of our early efforts in “translating research into practice” jumped over these steps.

Studying Implementation



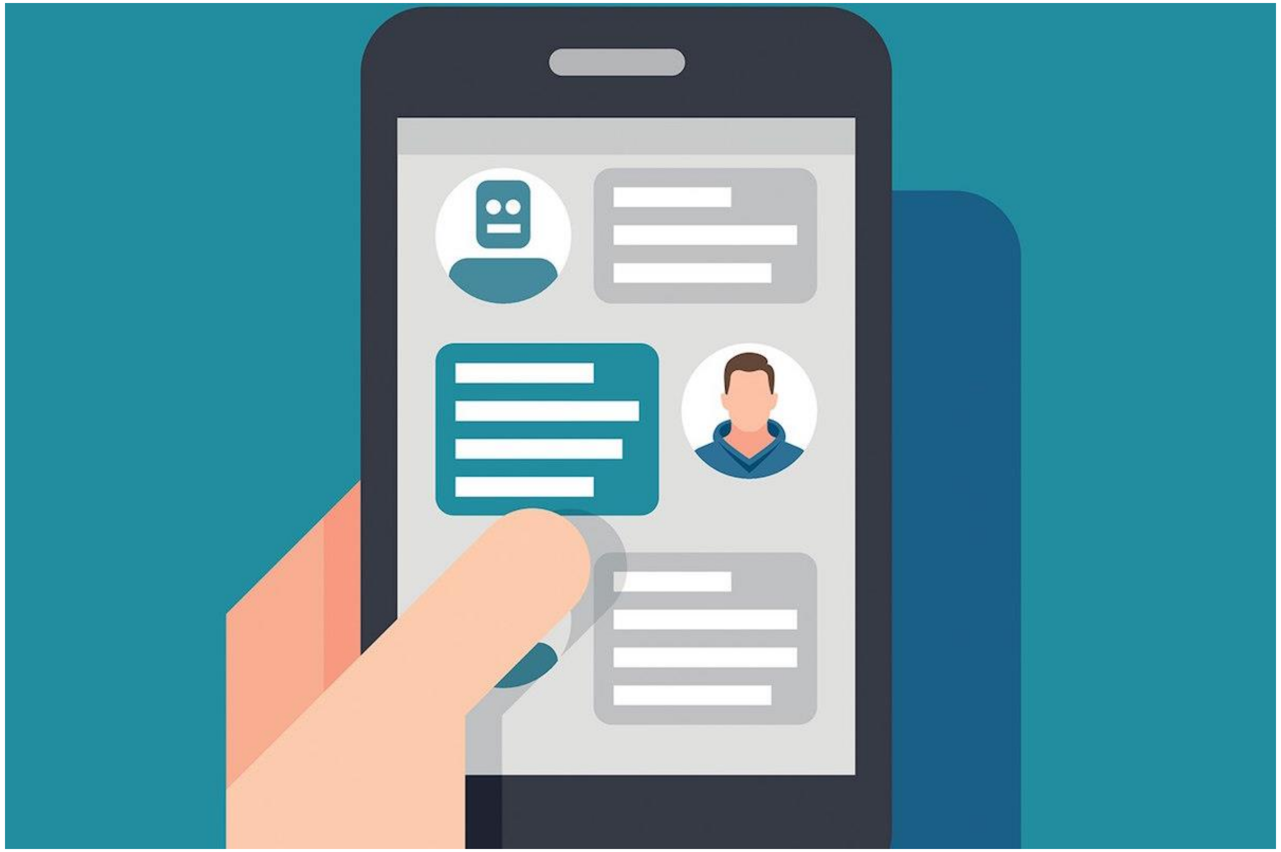
Example: Lung Cancer Screening



Sample IS Challenges:

- Is lung cancer screening a priority?
- How to reach all patients who could benefit
- Fit with practice workflow
- Implementing the model across varied practices
- How to bill for it?
- Workforce capacity/training needs

Example: AI-Driven Chat Bots



Source: https://blog.intakeq.com/can-healthcare-chatbots-improve-the-patient-experience/

- How does the chatbot fit into ongoing workflow?
- What is the start and end of the tech use?
- What else is present in the system to optimize benefit of chatbots?
- IT capabilities?
- Workforce training?
- How is the technology updated over time?

The Importance of What…

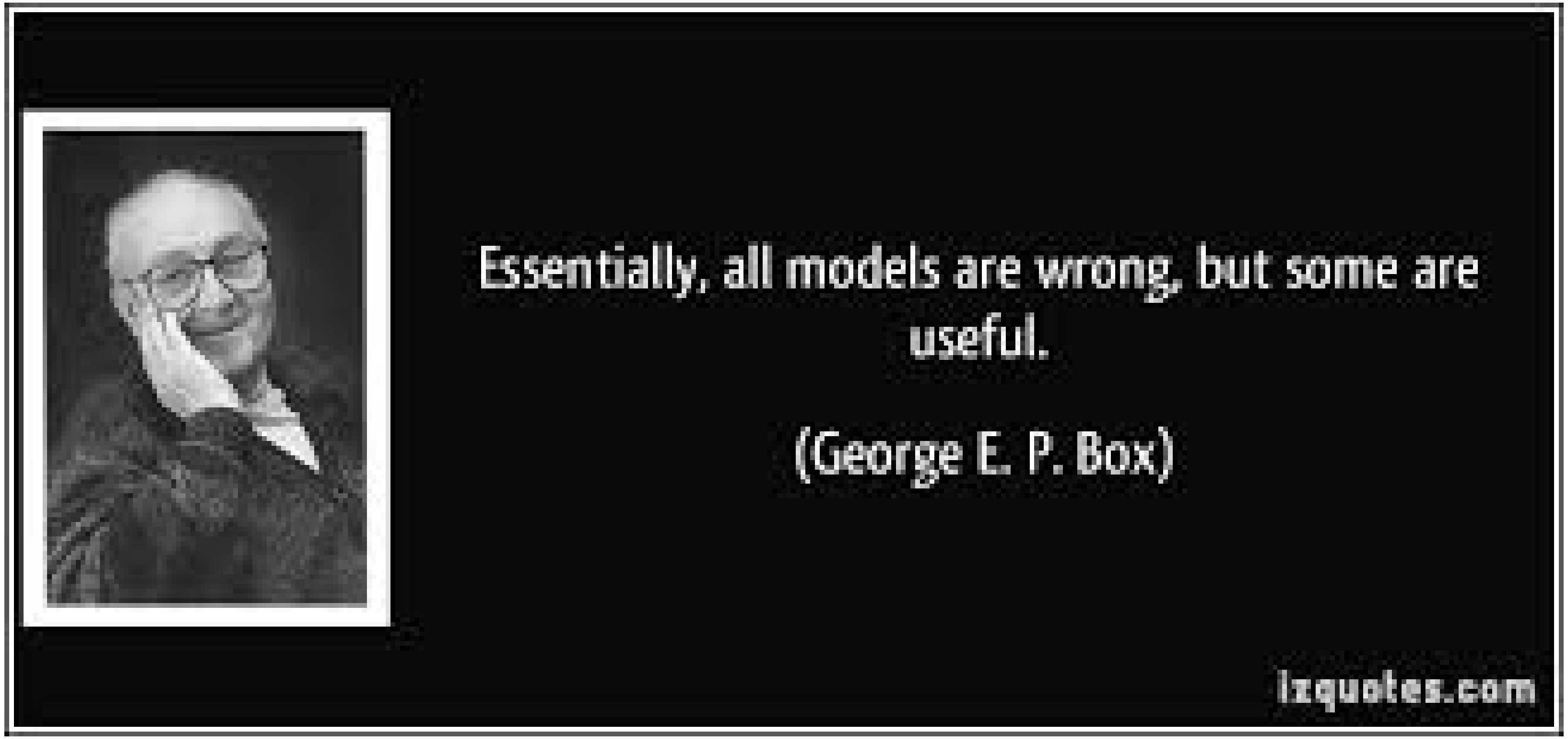
What is the intervention that needs to be implemented?

A. Diagnostic tests
B. Information Dissemination/educational intervention
C. Preventive Care
D. Treatment
E. Integrated Care
F. All of the above?

The fish-bicycle conundrum…

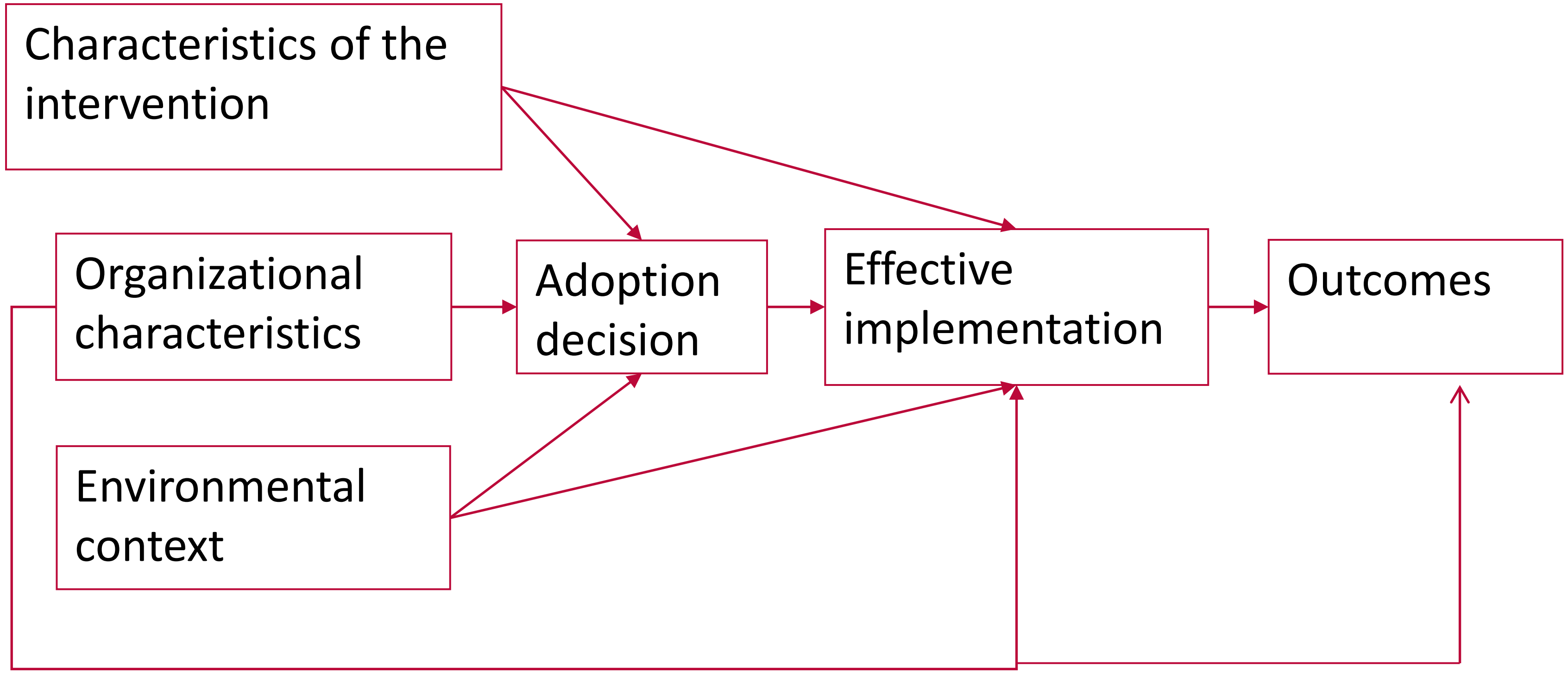
Ref: Paraphrased from Irina Dunn, 1970

IS Models, Theories and Frameworks



Ref: Tabak, Khoong, Chambers, Brownson, 2012, *AJPM*
http://www.dissemination-implementation.org

Roger's Diffusion of Innovations



CIPRS: Stetler & Damschroder Theoretical Frameworks

Krein SL, Olmsted RN, Hofer TP, Kowalski C, Forman J, Banaszak-Holl J, et al. Translating infection prevention evidence into practice using quantitative and qualitative research. Am. J. Infect. Control 2006;34(8):507-12.

EXPLORATION

OUTER CONTEXT

Sociopolitical Context
- Legislation
- Policies
- Monitoring and review

Funding
- Service grants
- Research grants
- Foundation grants
- Continuity of funding

Client Advocacy
- Consumer organizations

Interorganizational networks
- Direct networking
- Indirect networking
- Professional organizations
- Clearinghouses
- Technical assistance centers

INNER CONTEXT

Organizational characteristics
- Absorptive capacity
  - Knowledge/skills
  - Readiness for change
  - Receptive context
- Culture
- Climate
- Leadership

Individual adopter characteristics
- Values
- Goals
- Social Networks
- Perceived need for change

ADOPTION DECISION / PREPARATION

OUTER CONTEXT

Sociopolitical
- Federal legislation
- Local enactment
- Definitions of "evidence"

Funding
- Support tied to federal and state policies

Client advocacy
- National advocacy
- Class action lawsuits

Interorganizational networks
- Organizational linkages
- Leadership ties
- Information transmission
  - Formal
  - Informal

INNER CONTEXT

Organizational characteristics
- Size
- Role specialization
- Knowledge/skills/expertise
- Values

Leadership
- Culture embedding
- Championing adoption

ACTIVE IMPLEMENTATION

OUTER CONTEXT

Sociopolitical
- Legislative priorities
- Administrative costs

Funding
- Training
- Sustained fiscal support
- Contracting arrangements
- Community based organizations.

Interorganizational networks
- Professional associations
- Cross-sector
- Contractor associations
- Information sharing
- Cross discipline translation

Intervention developers
- Engagement in implementation

Leadership
- Cross level congruence
- Effective leadership practices

INNER CONTEXT

Organizational Characteristics
- Structure
- Priorities/goals
- Readiness for change
- Receptive context
- Culture/climate

Innovation-values fit
- EBP structural fit
- EBP ideological fit

Individual adopter characteristics
- Demographics
- Adaptability
- Attitudes toward EBP

SUSTAINMENT

OUTER CONTEXT

Sociopolitical
- Leadership
- Policies
  - Federal initiatives
  - State initiatives
  - Local service system
  - Consent decrees

Funding
- Fit with existing service funds
- Cost absorptive capacity
- Workforce stability impacts

Public-academic collaboration
- Ongoing positive relationships
- Valuing multiple perspectives

INNER CONTEXT

Organizational characteristics
- Leadership
- Embedded EBP culture
- Critical mass of EBP provision
- Social network support

Fidelity monitoring/support
- EBP Role clarity
- Fidelity support system
- Supportive coaching

Staffing
- Staff selection criteria
- Validated selection procedures

Aarons, G.A., Hurlburt, M. & Horwitz, S.M. (2011). Advancing a Conceptual Model of Evidence-Based Practice Implementation in Public Service Sectors. *Administration and Policy in Mental Health and Mental Health Services Research.38*, 4-23.

NIH Funding Opportunities

- R01: **PAR-19-274**
- R21: **PAR-19-275**
- R03: **PAR-19-276**

Dissemination and Implementation Research in Health

(R01/R21/R03)

- 21 NIH Institutes, Centers, and Offices
  - FIC, NCI, NHLBI, NHGRI, NIA, NIAAA, NIAID, NIAMS, NICHD, NIDCD, NIDCR, NIDA, NIEHS, NIMH, NINDS, NINR, NIMHD, NCCIH, ODP, OBSSR, ORWH

- Standing Study Section (Science of Implementation in Health and Healthcare (SIHH))

- >300 studies funded since the first round of the PARs

Select NCI-Funded IS Grants

R01: Implementing Tobacco Use Treatment Guidelines in Community Health Centers in Vietnam

Principal Investigator

Donna Shelley, MD, MPH
NEW YORK UNIVERSITY*

FOA**
PAR 13-055
Award Number
R01#CA175329-01A1

View Funded Grant (PDF, 2.21MB)

R01: De-implementation of low value castration for men with prostate cancer

Principal Investigator

Ted Skolarus, MD, MPH, FACS
UNIVERSITY OF MICHIGAN AT ANN ARBOR*

FOA**
PAR 16-238
Award Number
R37#CA222885-01

View Funded Grant (PDF, 828.73KB)

R01: Disseminating an Evidence-Based Tobacco Control Intervention for School Teachers in India

Principal Investigator

Glorian Sorenson, PhD, MPH
HARVARD SCHOOL OF PUBLIC HEALTH*

FOA**
PAR 13-055
Award Number
R01#CA200691-01A1

View Funded Grant (PDF, 1.26MB)

R21: Effective Training Models for Implementing Health-Promoting Practices Afterschool

Principal Investigator

Rebekka Mairghread Lee, ScD
HARVARD SCHOOL OF PUBLIC HEALTH*

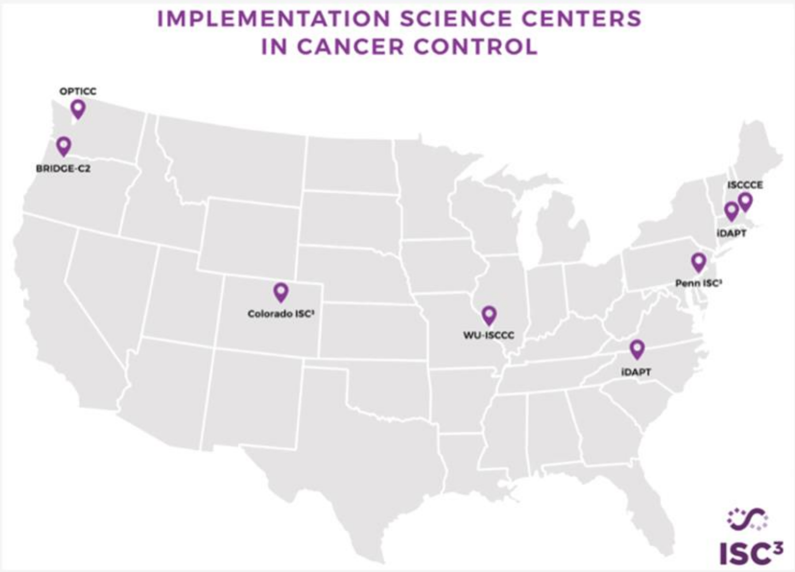
FOA**
PAR 13-054
Award Number
R21#CA201567-01A1

View Funded Grant (PDF, 836.74KB)

https://cancercontrol.cancer.gov/is/funding/sample-grant-applications

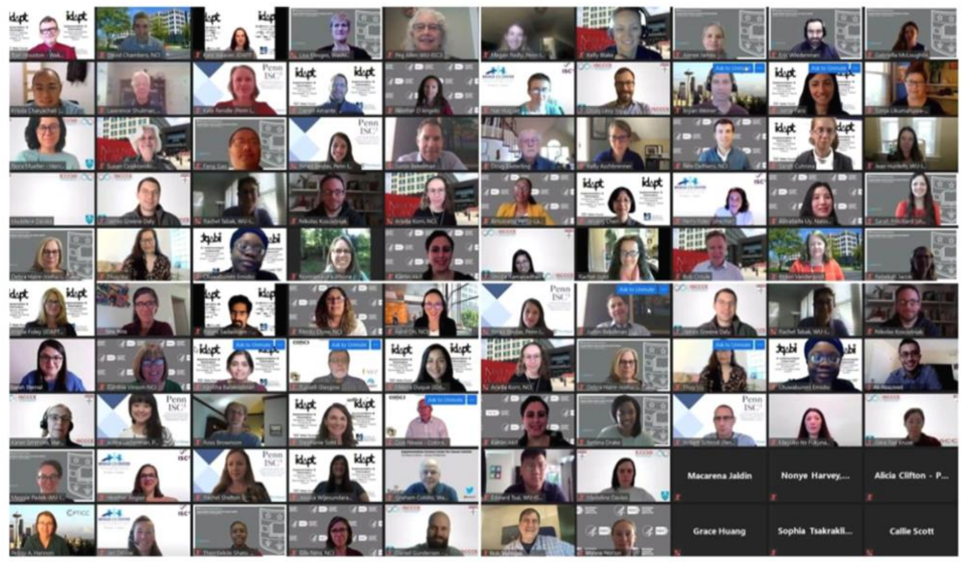
Implementation Science Centers in Cancer Control (ISC^3)

The ISC^3 Program is composed of six Centers funded by RFA-CA-19-005 and RFA-CA-19-006.



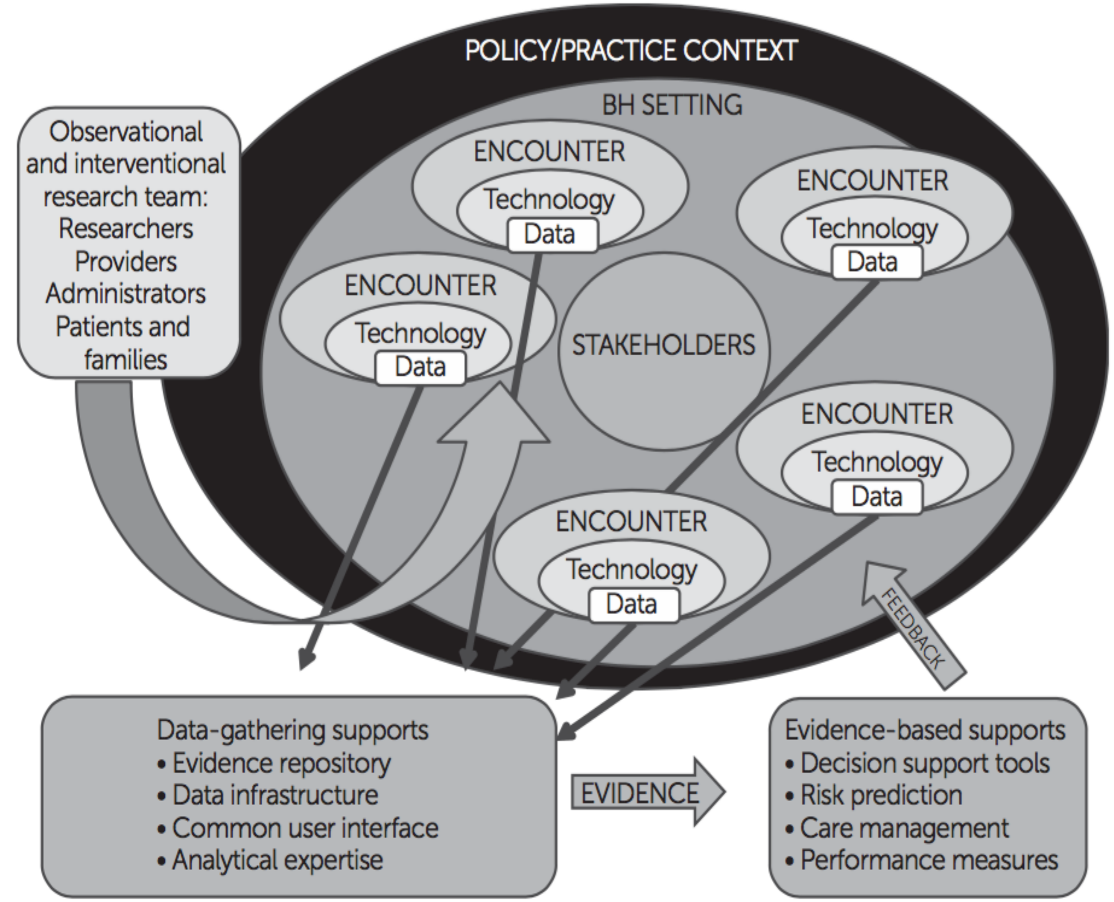
Center	PI/MPI	Institution
The Implementation Science Center for Cancer Control Equity (ISCCCE)	Karen Emmons Elsie Taveras	Harvard T.H. Chan School of Public Health
Building Research in Implementation and Dissemination to close Gaps and achieve Equity in Cancer Control Center (BRIDGE-C2)	Jennifer DeVoe Heather Angier Nathalie Huguet	Oregon Health & Science University
Colorado Implementation Science Center in Cancer Control (Colorado ISC^3)	Russell E. Glasgow	University of Colorado School of Medicine
Optimizing Implementation in Cancer Control (OPTICC)	Bryan J. Weiner Margaret Hannon Cara C. Lewis	University of Washington
Implementation and Informatics – Developing Adaptable Processes and Technologies for Cancer Control (iDAPT)	Kristie Long Foley Thomas Houston Sarah Cutrona	Wake Forest School of Medicine/University of Massachusetts Medical School
Washington University Implementation Science Center for Cancer Control (WU-ISCCC)	Ross C. Brownson Graham A. Colditz	Washington University in St. Louis
* Penn Implementation Science Center in Cancer Control (Penn ISC^3)	Justin Bekelman Rinad Beidas Robert Schnoll	University of Pennsylvania

NCI Staff: Cynthia Vinson, April Oh (leads), Kelly Blake, Mindy Clyne, Robin Vanderpool, Amy Caplon, Heather D'angelo, Susan Czajkowski and more
https://cancercontrol.cancer.gov/IS/initiatives/ISC3.html



Ongoing Learning from Practice Settings

FIGURE 1. Collection and use of data to inform decision making by stakeholders in a learning behavioral health care (BH) system[a]



Stein, Adams, Chambers. *Psychiatric Services,* 2016.

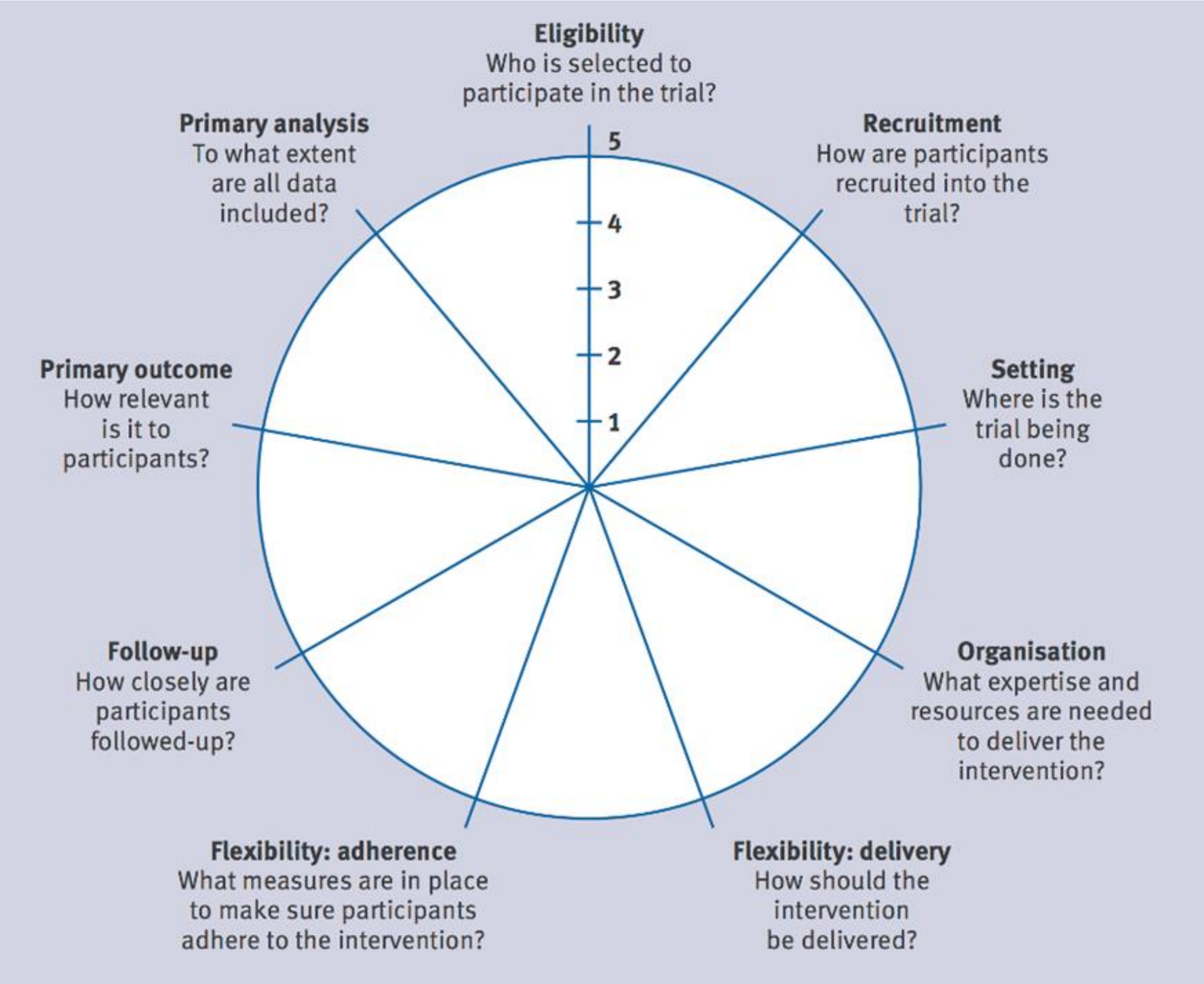
Moving Forward

https://jocatorres.medium.com/innovation-a-lot-of-opportunities-480be0d81f68

- **Study Design**
- **Exciting Areas**

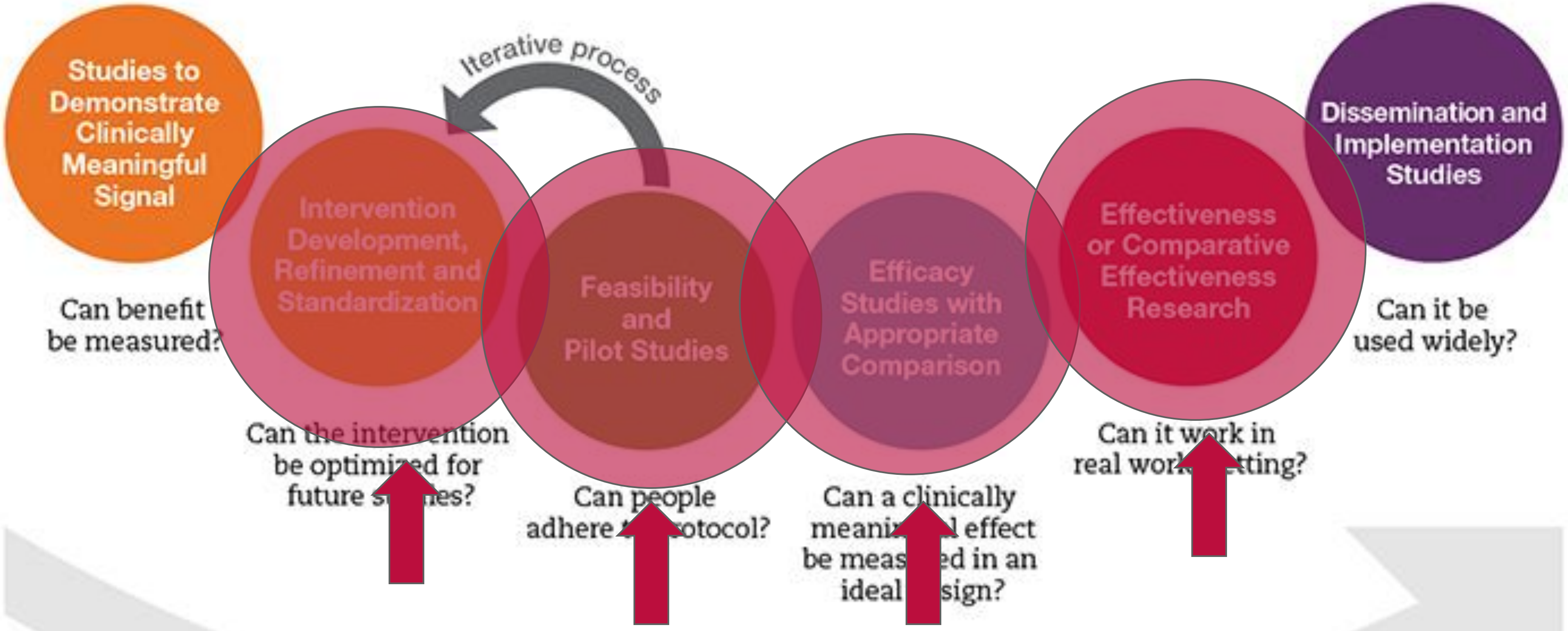
Reconsidering How we Design our Trials:

The PRagmatic-Explanatory Continuum Index Summary 2 (PRECIS-2) wheel



Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. **The PRECIS-2 tool: designing trials that are fit for purpose.** *BMJ.* 2015;350:h2147.

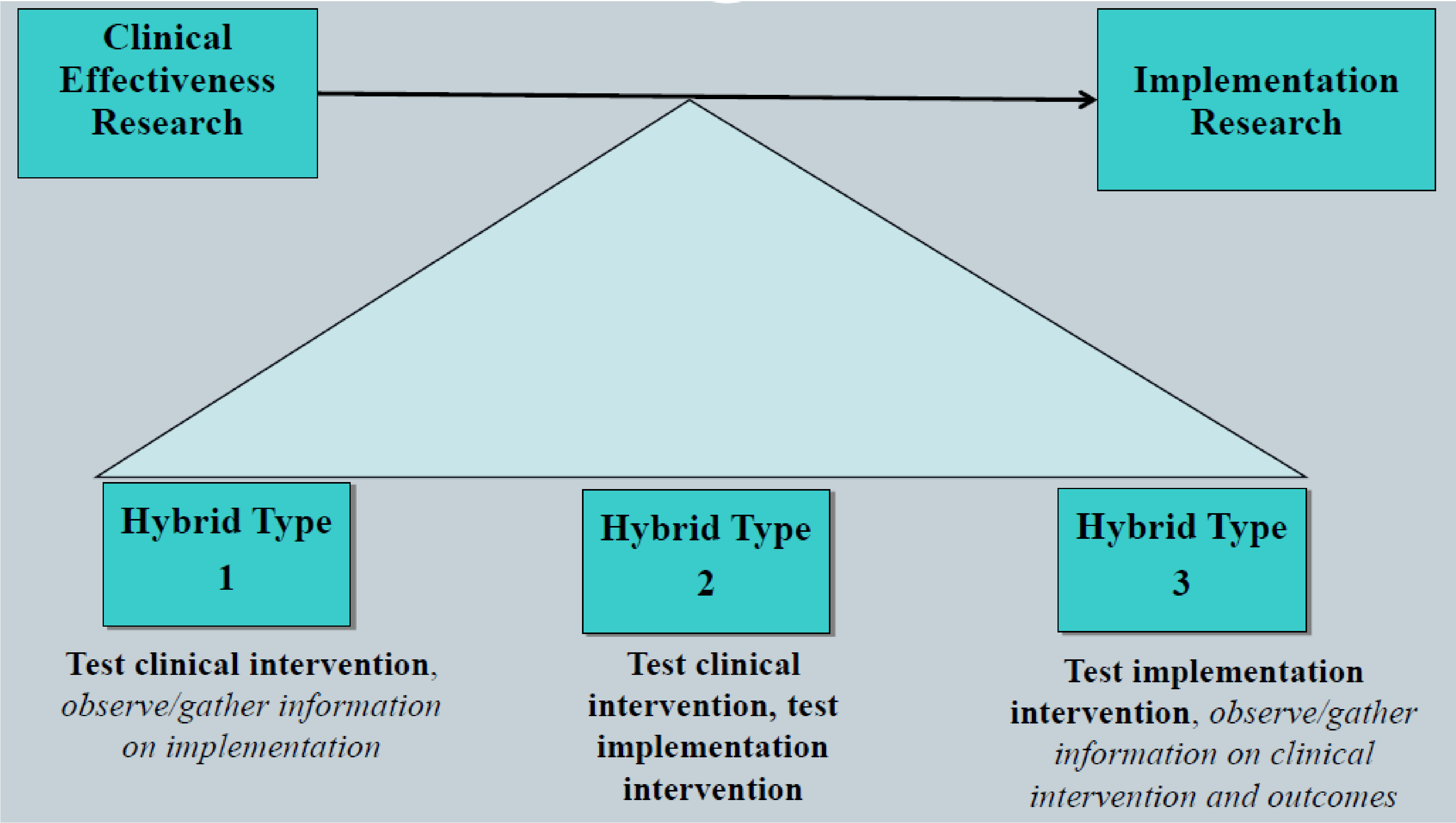
Considering D&I earlier



https://nccih

An earlier focus on…
- Who's going to deliver it?
- Fit with ultimate patient population
- Building in tests of training, support, adherence, mediators and moderators to high quality delivery
- Hybrid designs

Hybrid designs: 1, 2, 3



Curran et al. (2013). Effectiveness-implementation hybrid. Med Care.

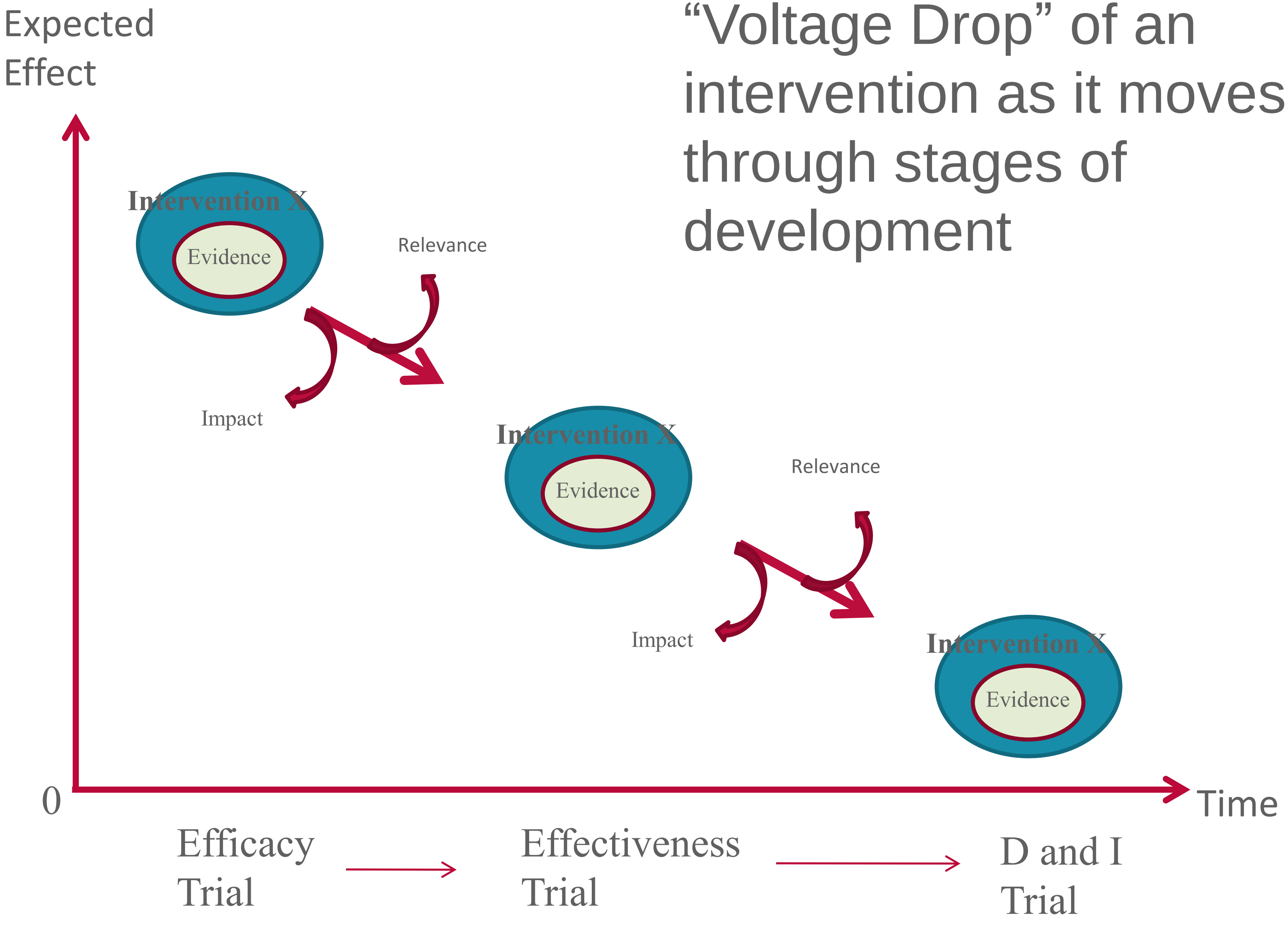
TOWARDS A DYNAMIC VIEW

Traditional Assumptions

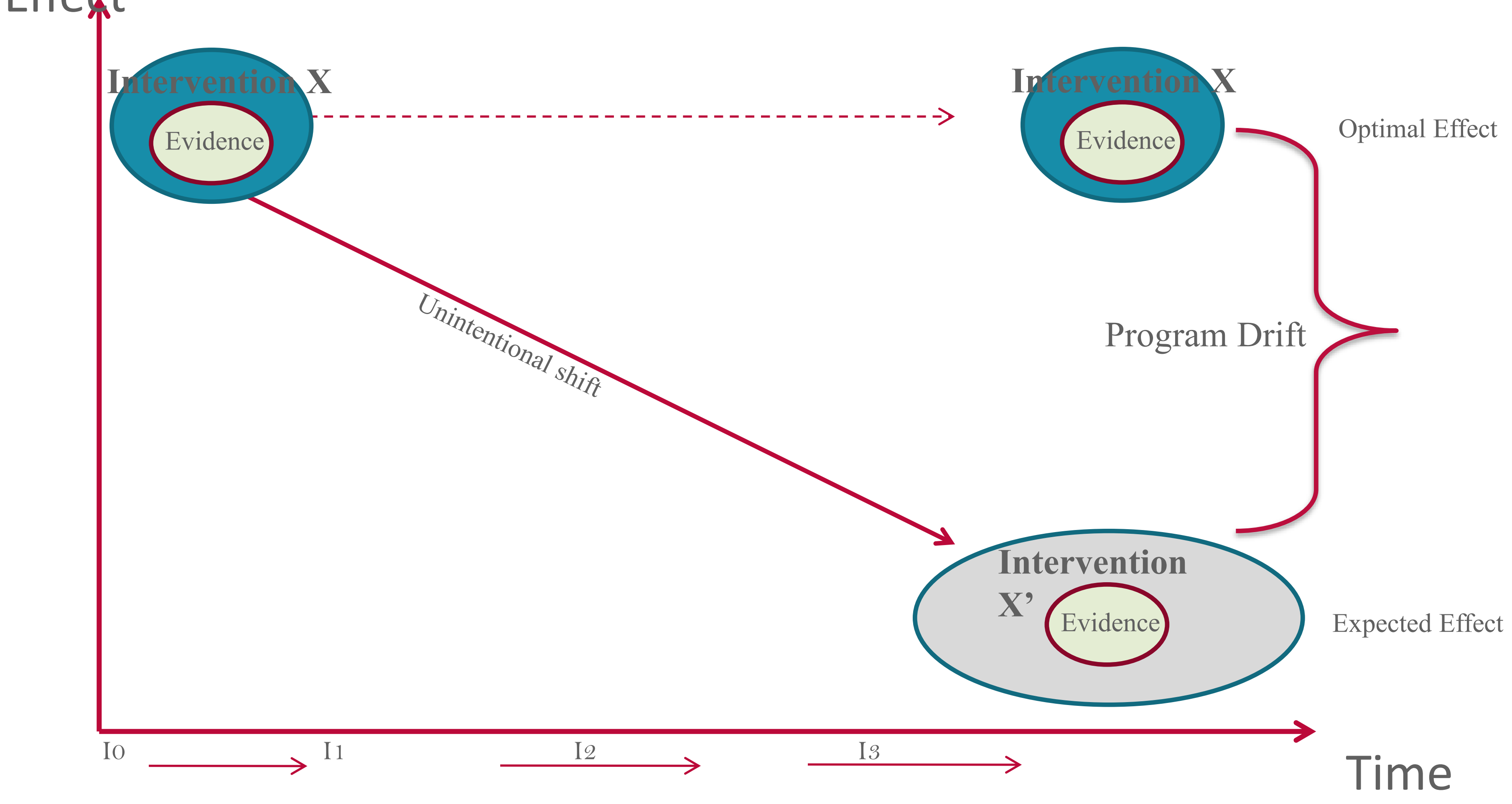
- Evidence-based Interventions are static
- System is static
- Implementation proceeds one practice or test at a time
- Consumers/Patients are homogeneous
- Choosing to not implement is irrational

Chambers, Glasgow, Stange (2013), The Dynamic Sustainability Framework. *Implementation Science*

"Voltage Drop" of an intervention as it moves through stages of development

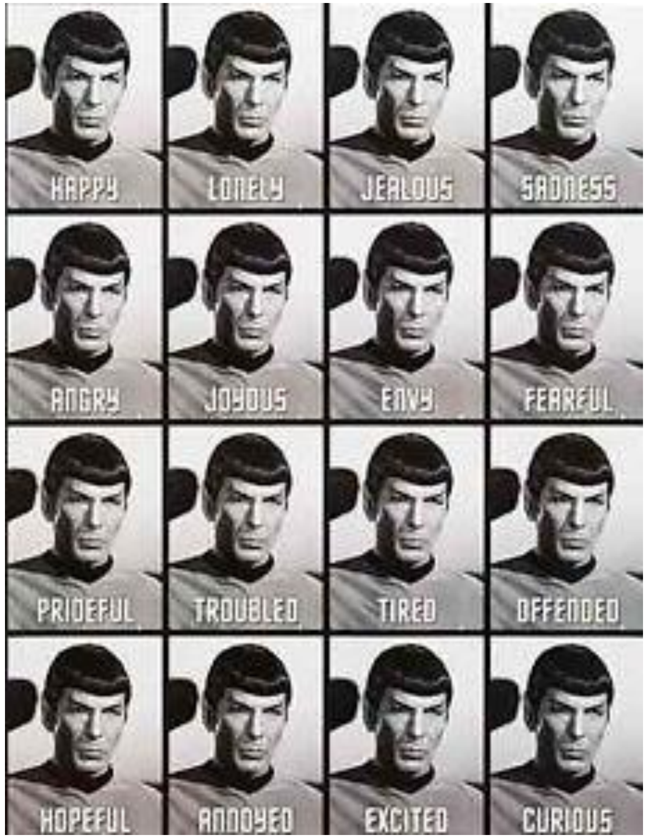


“Program Drift” of a fielded intervention (ITV) over time, with expected decrease of effect

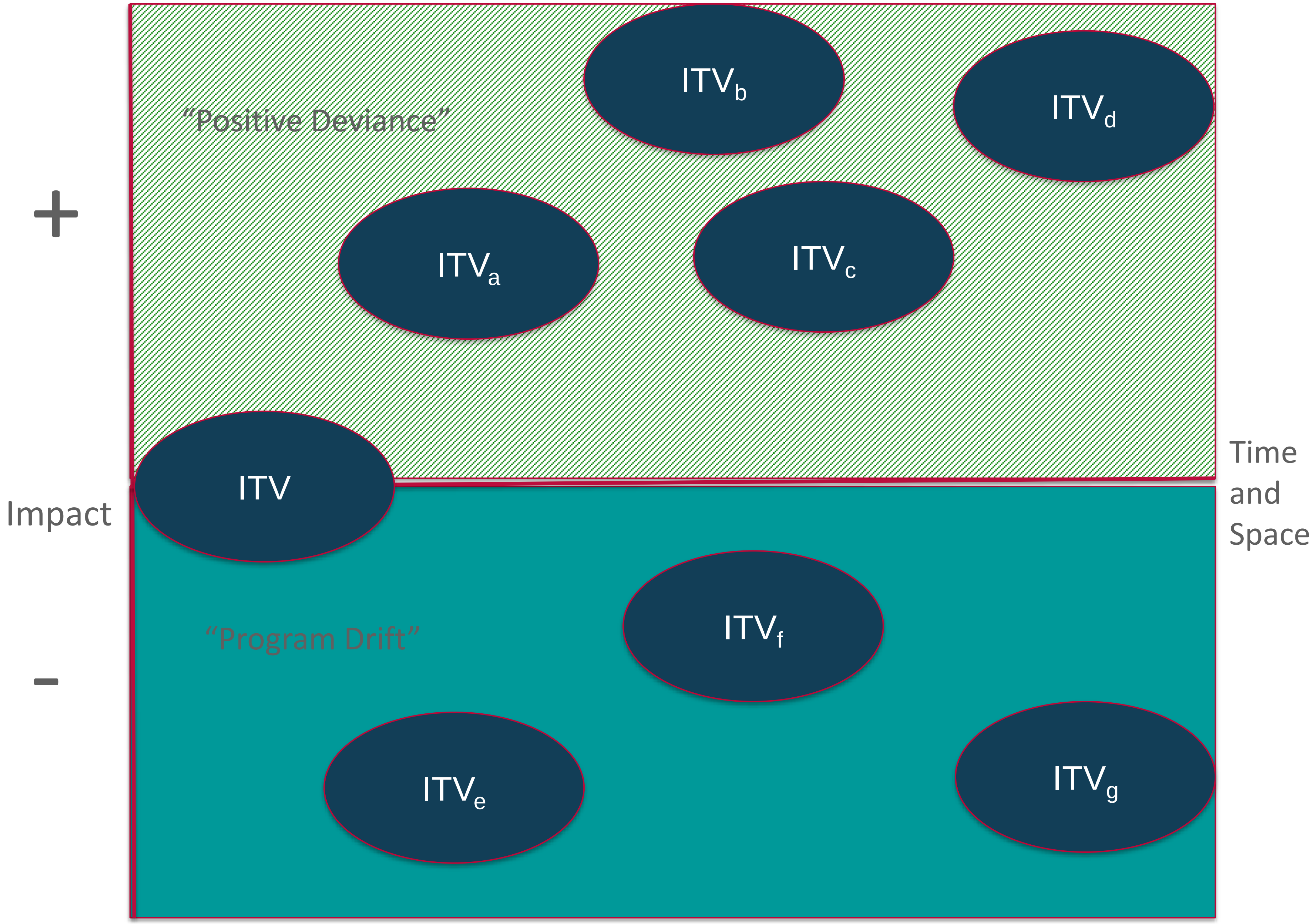


Chambers, Glasgow, Stange (2013), The Dynamic Sustainability Framework. *Implementation Science*

Fidelity vs Adaptation?



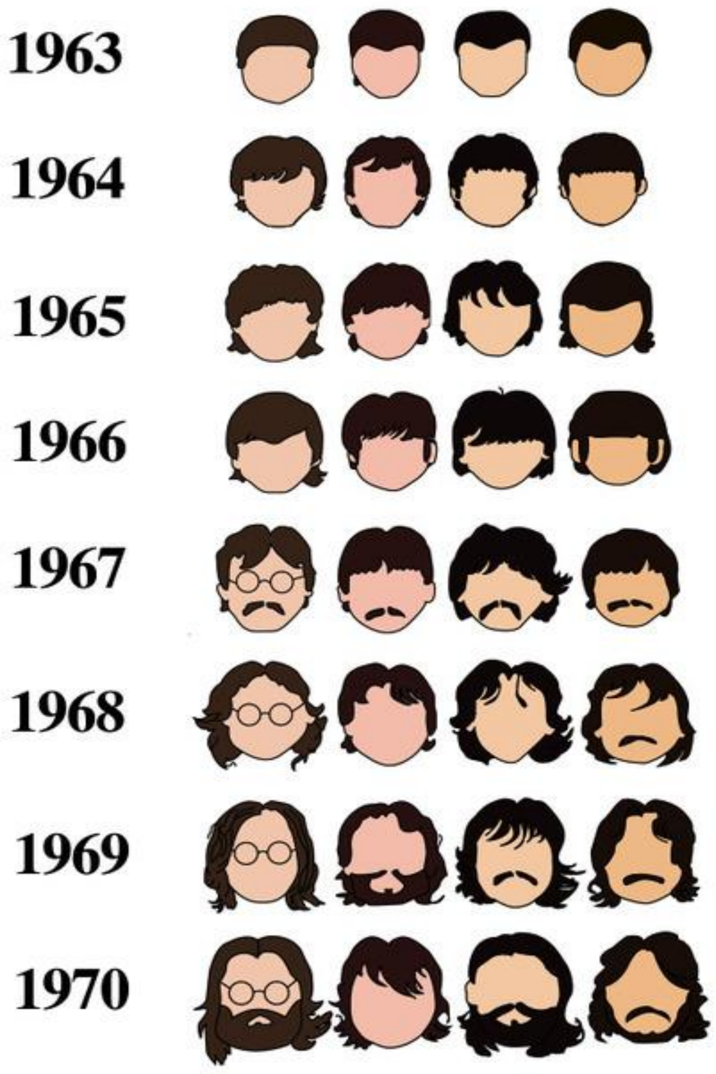
Variable use for variable populations, settings, and purposes…



KEY: ITV = Intervention, Time and Space = variability of intervention characteristics over time and setting

Embracing Dynamism

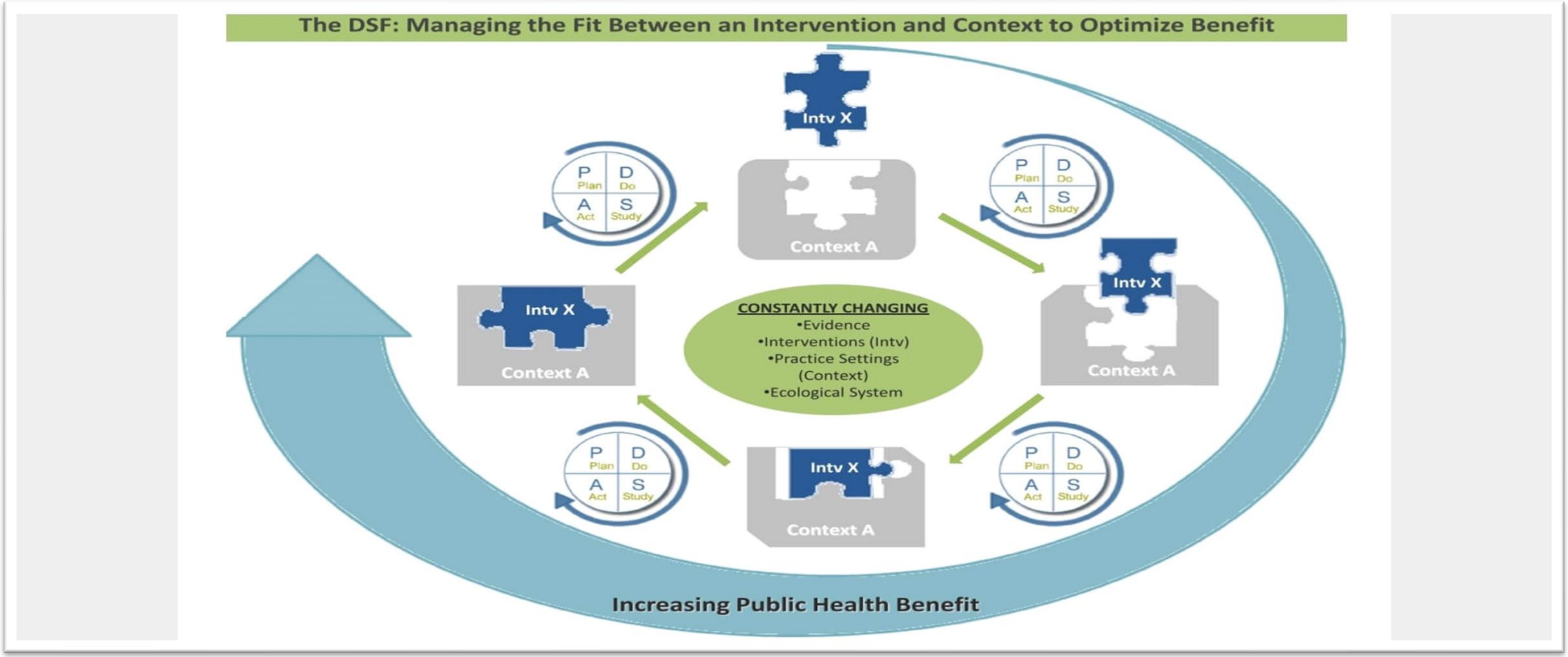
Sustainability or Evolution?



- IF MEDICINE CONTINUES TO EVOLVE, SHOULD EXISTING INTERVENTIONS BE SUSTAINED IN THE SAME FORM THAT WE'VE CREATED THEM?
- HOW DOES THE SYSTEM COPE WITH A DYNAMIC FIELD THAT IS CONSTANTLY CHANGING?
- WHERE DO WE GO FROM HERE?

http://www.thestrut.com/2012/12/19/the-evolution-of-the-beatles-hair/

A Dynamic Approach to Sustainability…

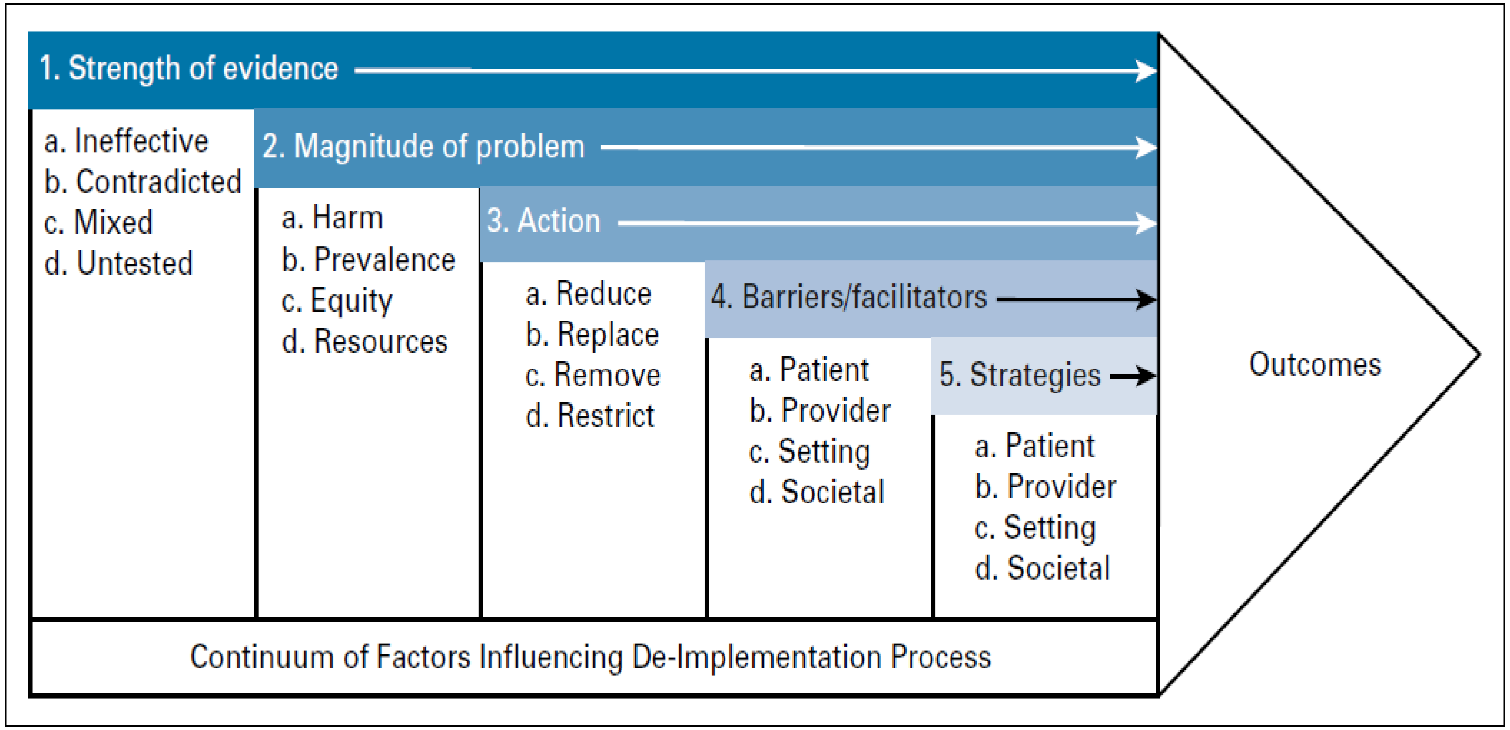


Chambers, Stange, & Glasgow, *Implementation Science*, 2013

SCALING UP INTERVENTIONS

What is it that we’re scaling up? Is it asked for? Can it be used?

Need: Understanding De-Implementation

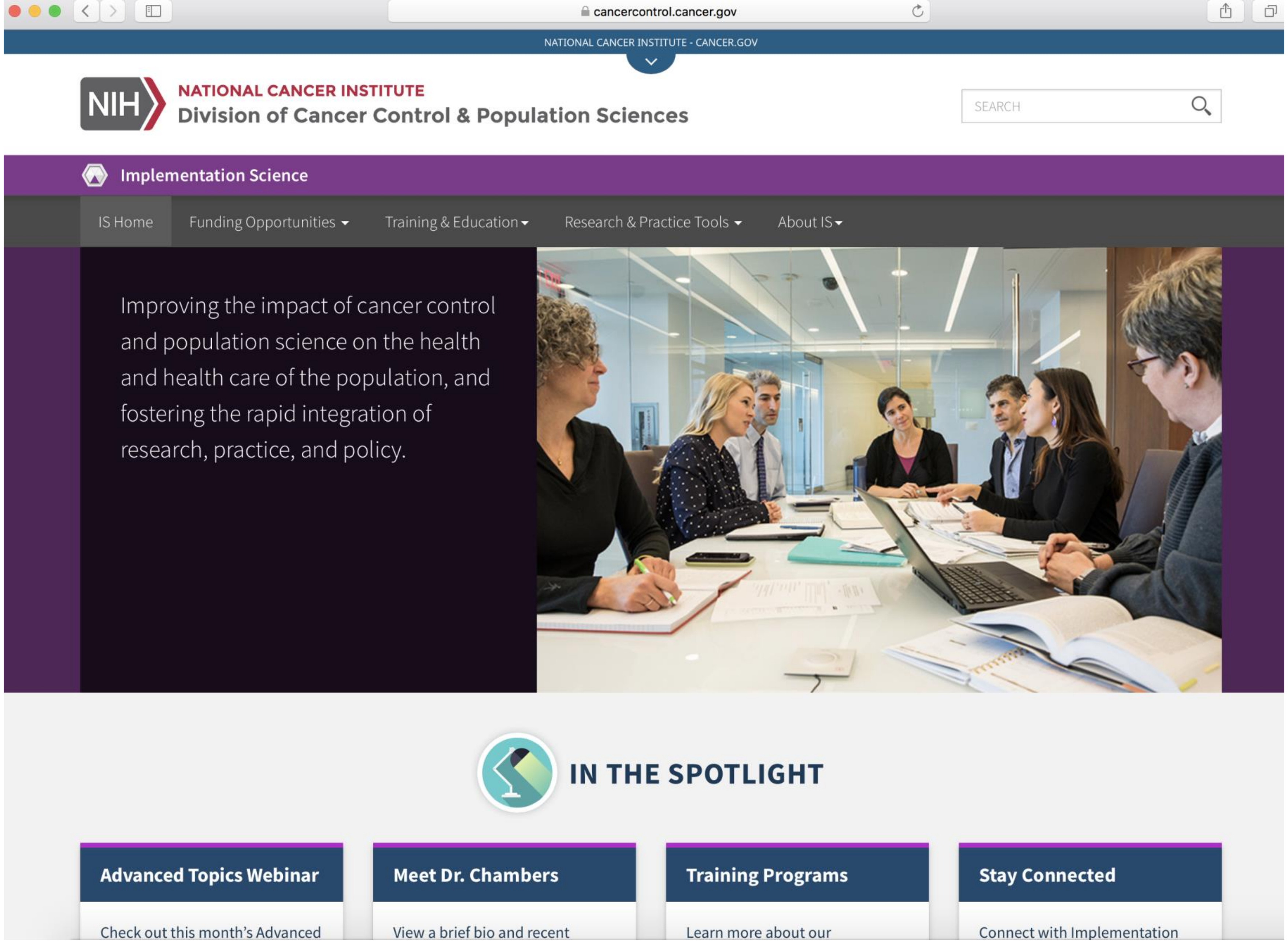


Norton, Chambers, & Kramer, *JCO*, 2018

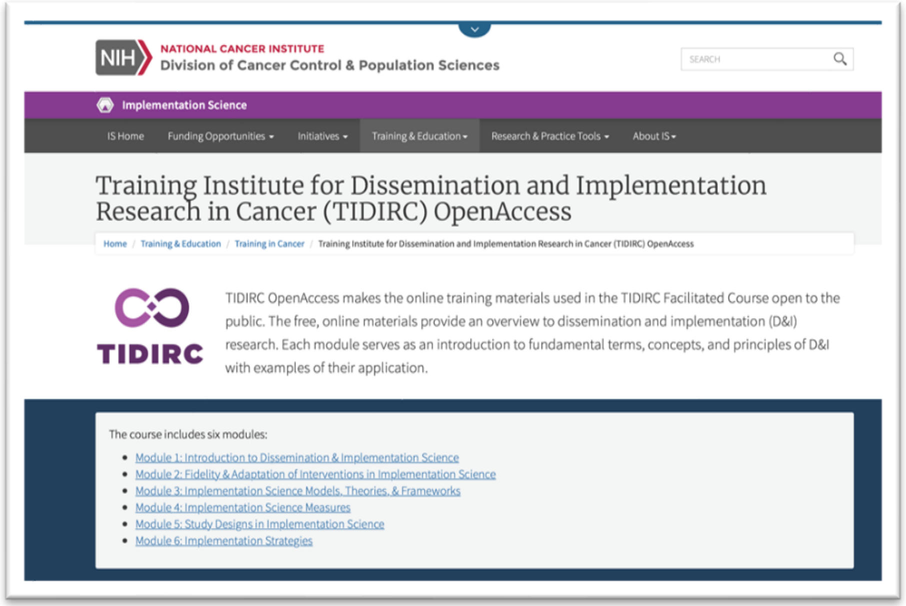
Selected Implementation Science Priorities

- Studies of the local adaptation of evidence-based practices in the context of implementation
- Longitudinal and follow-up studies on the factors that contribute to the sustainability of evidence-based interventions
- Scaling up health care interventions across health plans, systems, and networks
- De-Implementation of ineffective or suboptimal care

Connections to AI, Medical Imaging, Research/Policy/Practice?


cancercontrol.cancer.gov
NATIONAL CANCER INSTITUTE - CANCER.GOV
NIH
NATIONAL CANCER INSTITUTE
Division of Cancer Control & Population Sciences
SEARCH
Implementation Science
IS Home
Funding Opportunities
Training & Education
Research & Practice Tools
About IS
Improving the impact of cancer control and population science on the health and health care of the population, and fostering the rapid integration of research, practice, and policy.
IN THE SPOTLIGHT
Advanced Topics Webinar
Check out this month's Advanced
Meet Dr. Chambers
View a brief bio and recent
Training Programs
Learn more about our
Stay Connected
Connect with Implementation

Implementation Science Resources



NIH NATIONAL CANCER INSTITUTE Division of Cancer Control & Population Sciences

Implementation Science

IS Home | Funding Opportunities | Initiatives | Training & Education | Research & Practice Tools | About IS

Training Institute for Dissemination and Implementation Research in Cancer (TIDIRC) OpenAccess

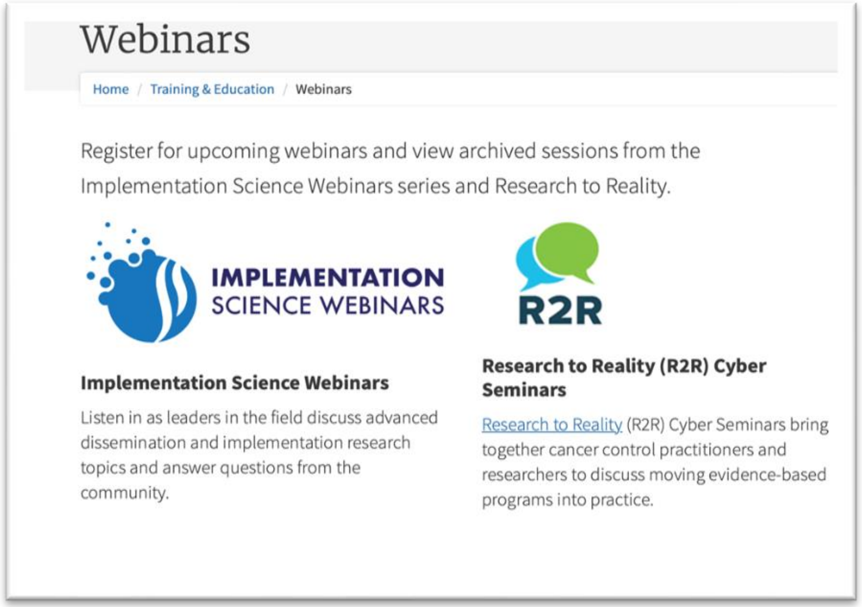
Home / Training & Education / Training in Cancer / Training Institute for Dissemination and Implementation Research in Cancer (TIDIRC) OpenAccess

TIDIRC

TIDIRC OpenAccess makes the online training materials used in the TIDIRC Facilitated Course open to the public. The free, online materials provide an overview to dissemination and implementation (D&I) research. Each module serves as an introduction to fundamental terms, concepts, and principles of D&I with examples of their application.

The course includes six modules:

- Module 1: Introduction to Dissemination & Implementation Science
- Module 2: Fidelity & Adaptation of Interventions in Implementation Science
- Module 3: Implementation Science Models, Theories, & Frameworks
- Module 4: Implementation Science Measures
- Module 5: Study Designs in Implementation Science
- Module 6: Implementation Strategies



Webinars

Home / Training & Education / Webinars

Register for upcoming webinars and view archived sessions from the Implementation Science Webinars series and Research to Reality.

IMPLEMENTATION SCIENCE WEBINARS

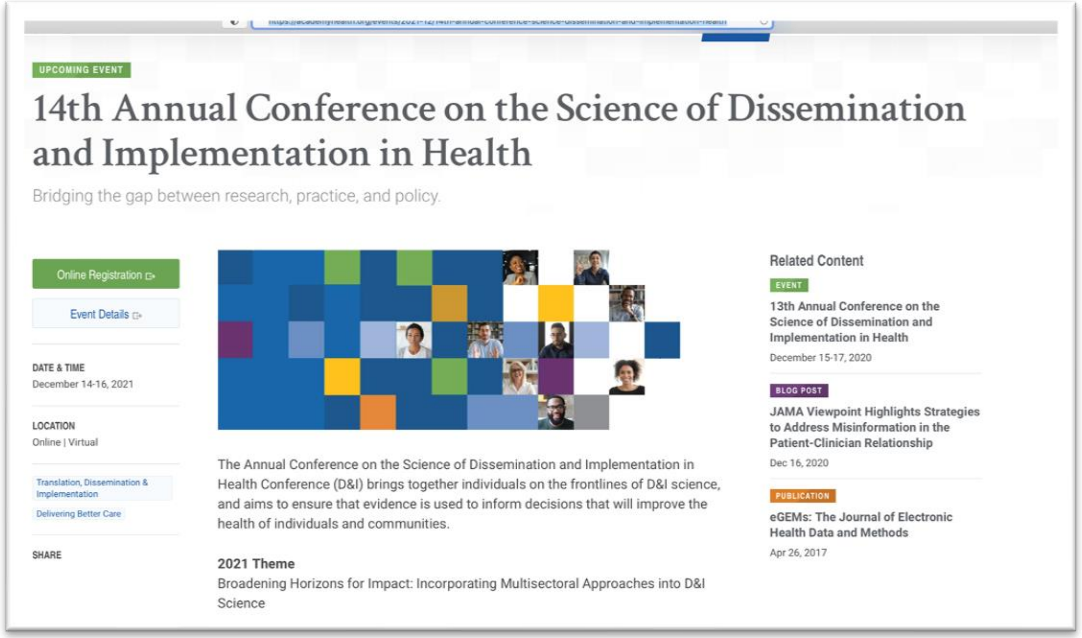
R2R

Implementation Science Webinars

Listen in as leaders in the field discuss advanced dissemination and implementation research topics and answer questions from the community.

Research to Reality (R2R) Cyber Seminars

Research to Reality (R2R) Cyber Seminars bring together cancer control practitioners and researchers to discuss moving evidence-based programs into practice.



UPCOMING EVENT

14th Annual Conference on the Science of Dissemination and Implementation in Health

Bridging the gap between research, practice, and policy.

Online Registration

Event Details

DATE & TIME
December 14-16, 2021

LOCATION
Online | Virtual

Translation, Dissemination & Implementation

Delivering Better Care

SHARE

The Annual Conference on the Science of Dissemination and Implementation in Health Conference (D&I) brings together individuals on the frontlines of D&I science, and aims to ensure that evidence is used to inform decisions that will improve the health of individuals and communities.

2021 Theme

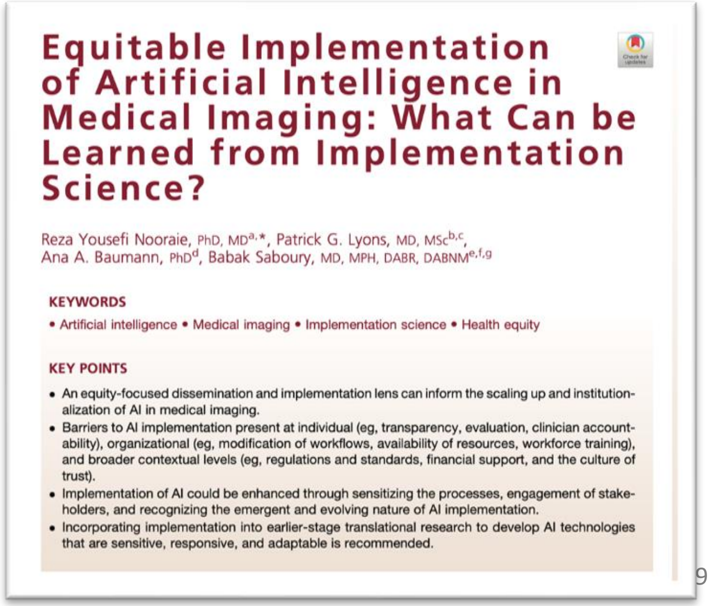
Broadening Horizons for Impact: Incorporating Multisectoral Approaches into D&I Science

Related Content

EVENT
13th Annual Conference on the Science of Dissemination and Implementation in Health
December 15-17, 2020

BLOG POST
JAMA Viewpoint Highlights Strategies to Address Misinformation in the Patient-Clinician Relationship
Dec 16, 2020

PUBLICATION
eGEMs: The Journal of Electronic Health Data and Methods
Apr 26, 2017



Equitable Implementation of Artificial Intelligence in Medical Imaging: What Can be Learned from Implementation Science?

Reza Yousefi Nooraie, PhD, MD[a,*], Patrick G. Lyons, MD, MSc[b,c], Ana A. Baumann, PhD[d], Babak Saboury, MD, MPH, DABR, DABNM[e,f,g]

KEYWORDS

- Artificial intelligence • Medical imaging • Implementation science • Health equity

KEY POINTS

- An equity-focused dissemination and implementation lens can inform the scaling up and institutionalization of AI in medical imaging.
- Barriers to AI implementation present at individual (eg, transparency, evaluation, clinician accountability), organizational (eg, modification of workflows, availability of resources, workforce training), and broader contextual levels (eg, regulations and standards, financial support, and the culture of trust).
- Implementation of AI could be enhanced through sensitizing the processes, engagement of stakeholders, and recognizing the emergent and evolving nature of AI implementation.
- Incorporating implementation into earlier-stage translational research to develop AI technologies that are sensitive, responsive, and adaptable is recommended.

- **Consortium for Cancer Implementation Science** (CCIS)
  - First meeting: July 10-12, 2019, 243 participants (in-person and online)
  - Second meeting: Sept 22-23, 2020, 411 participants (online)
  - Third meeting: October 6-7, 2021, 800 registrants (online), ***Re-emergence from the Pandemic: Implementing Lessons Learned and Moving Ahead***
- **Action Teams** developing "public goods" in: equity and context, learning health systems, global IS, policy, community participation, technology, multi-level interventions, and study designs; Utilizing **SLACK** platform
- *Small contract opportunities for key action group products
- ***Fall/Winter NCI IS Webinars** will feature 3 CCIS topics

https://cancercontrol.cancer.gov/IS/initiatives/ccis.html

dchamber@mail.nih.gov
240-276-5090
@NCIDAChambers