Welcome to the ISPOR ***Signal*** Series Episode

Larger, Deeper and in Real Time: Applications of Machine Learning and Natural Language Processing on Electronic Health Records

#ISPORSignal

Exploring what will shape healthcare decision making over the next decade...

#ISPORSignal

Larger, Deeper and in Real Time: Applications of Machine Learning and Natural Language Processing on Electronic Health Records

#ISPORSignal

Larger, Deeper, and in Real Time:

Applications of Machine Learning and Natural Language Processing on Electronic Health Records to Learn from the Patient Journey at Scale

Discussion Leader:

Joe Vandigo, MBA, PhD

Discussants:

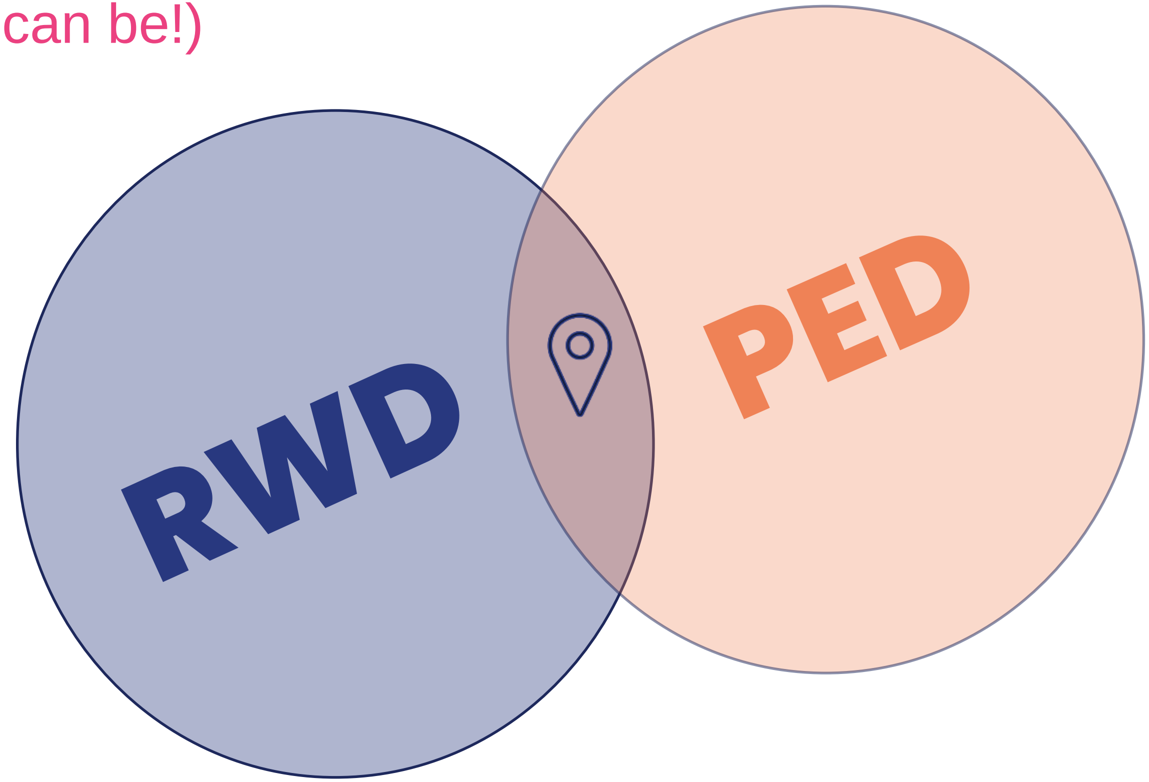
Selen Bozkurt, PhD, MS

Katherine Tan, PhD

Ravi Parikh, MD, MPP

RWD is not always Patient Experience Data

(But it can be!)



Improved mapping of the patient experience has implications for RWD study designs.

First interaction with healthcare system

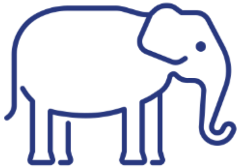
- Target Population
- Hypothesis Generating
- Study Period

Treatments and side effects

- Exposure
- Potential Outcomes
- Covariates & Confounders

SOURCE: Oehrlein EM, Burcu M, Schoch S, Gressler LE. **Enhancing Patient Centricity of Real-World Data Research: An Exploratory Analysis Using the Patient Experience Mapping Toolbox**. *Value Health*. 2023;26(1):10-17. doi:10.1016/j.jval.2022.10.002

Themes for today's panel



Scale

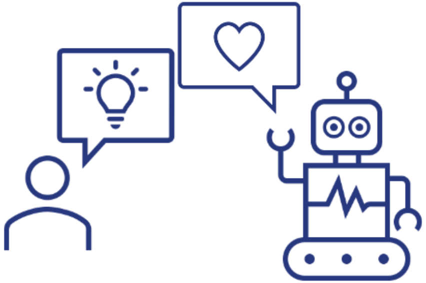
Increasing stratification cohort sizes

Speed

Keeping up with standard of care

Depth

Improving representation of underserved populations



Human and AI Collaboration

Key Questions

What is the most exciting opportunity in capturing the patient journey - scale, speed, or depth?

How and when can patients and other stakeholders engage in ML/NLP processes?

Panelists

Joe Vandigo, MBA, PHD

Moderator
Applied Patient Experience

Selen Bozkurt, PhD, MS

Discussant
Stanford University

Katherine Tan, PhD

Discussant
Flatiron Health

Ravi Parikh, MD, MPP

Discussant
University of Pennsylvania

Unlocking the Power of Electronic Health Records with NLP/ML

Selen Bozkurt, PhD, MS

Senior Research/Data Scientist

Stanford University School of Medicine (Biomedical Informatics)

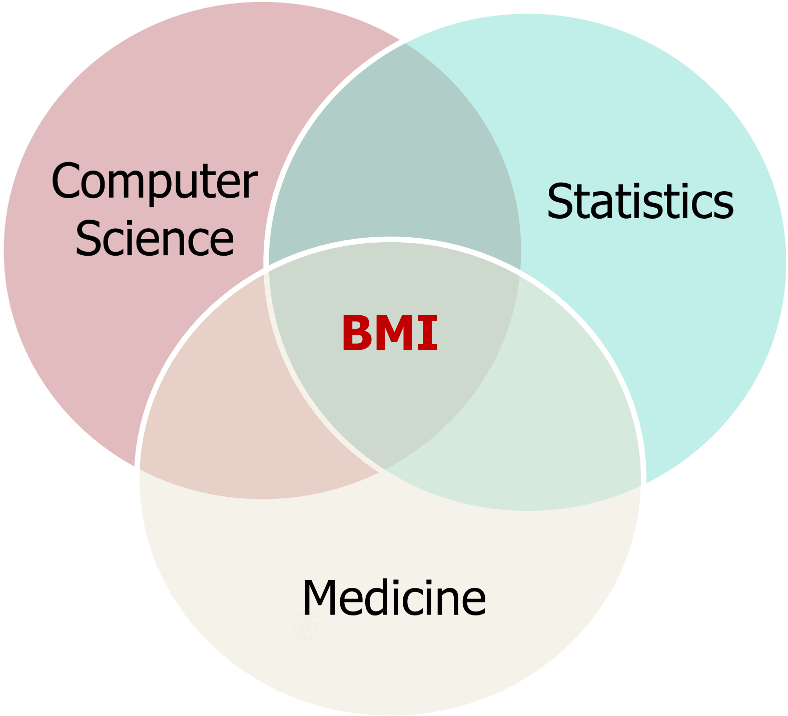
VA Palo Alto, Center for Innovation to Implementation

Incoming Assistant Professor, Emory University, Faculty of Medicine (Biomedical Informatics)

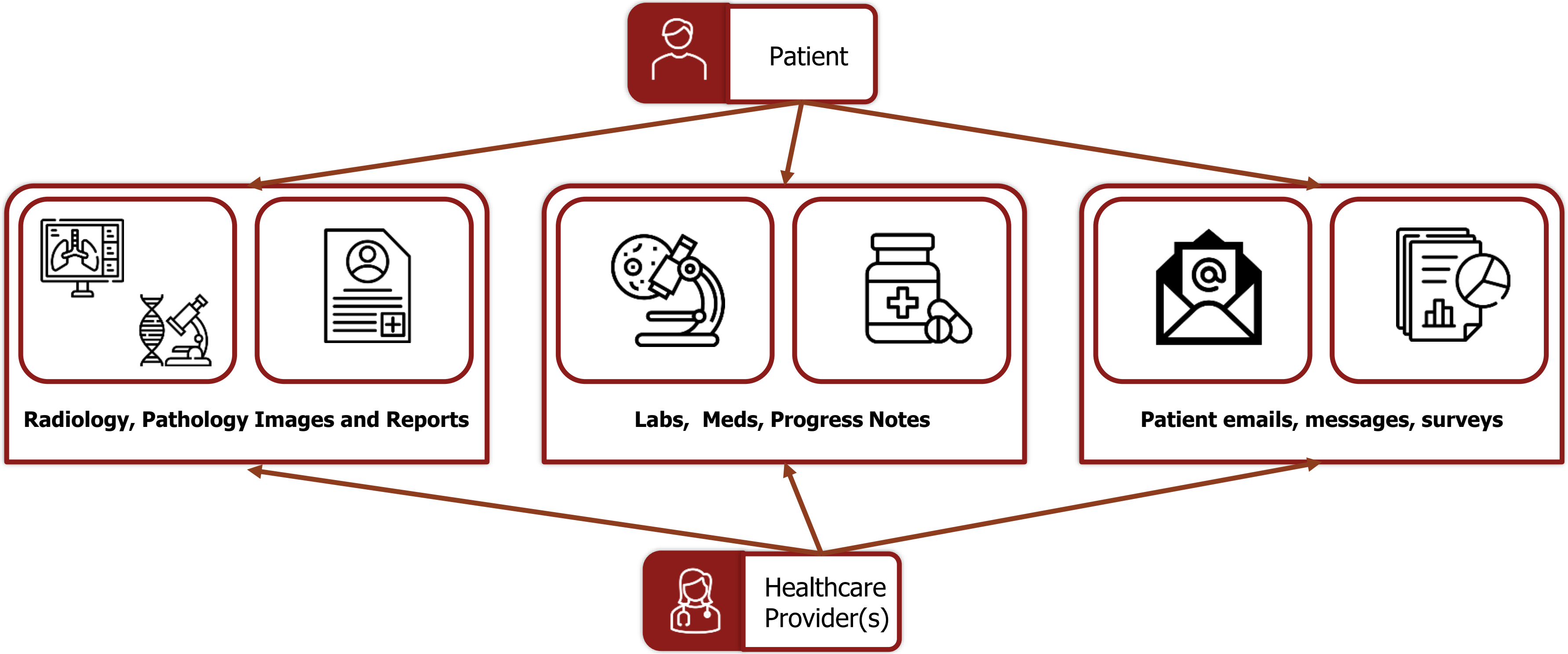
Disclosures

Academic consultancy services to F2IL, Flatiron Health, an independent subsidiary of the Roche Group.

Biomedical Informatics

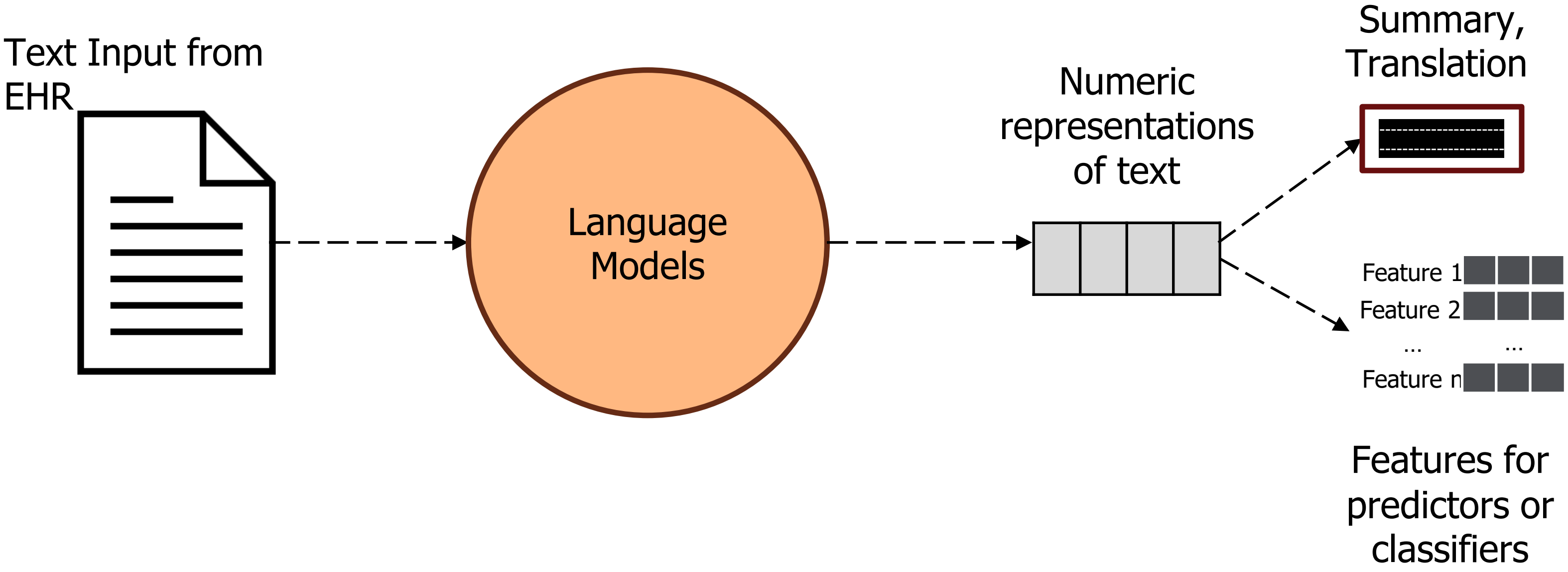


Patient journey through EHR documentation

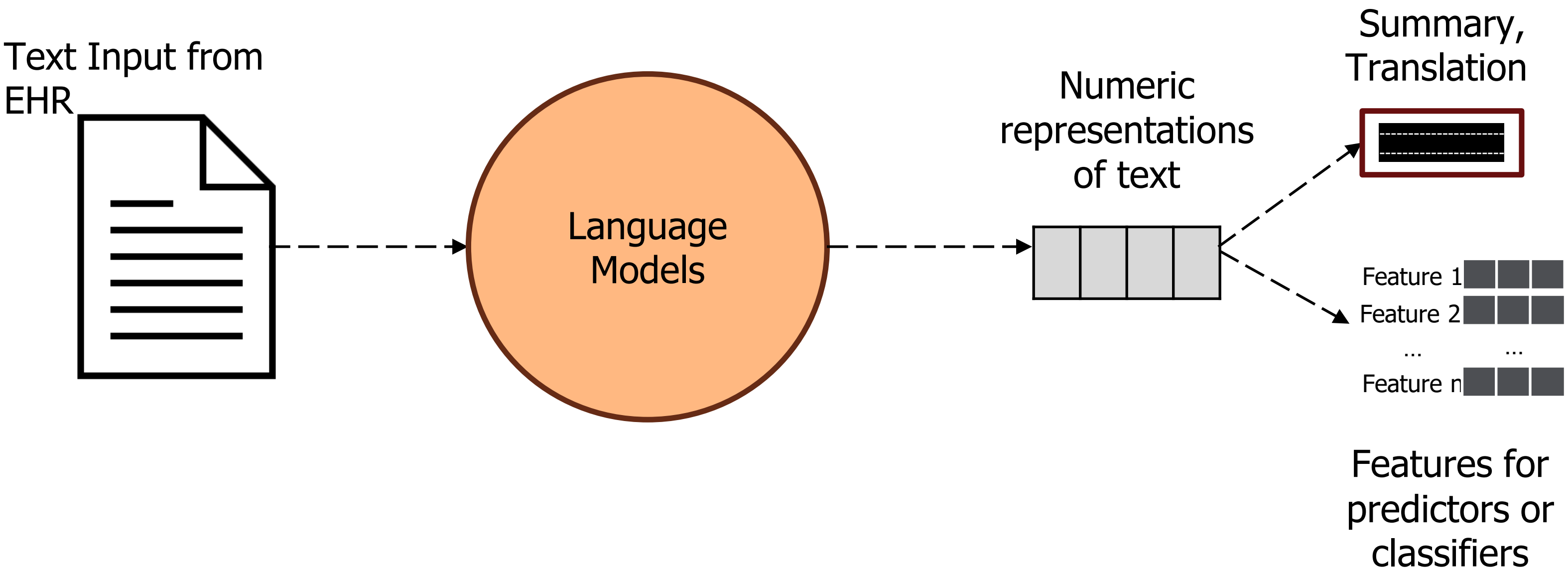


Longitudinally & Multi Specialty

Learning from (clinical) text



Learning from (clinical) text



Which text?

Which language?
Which model?

How to summarize?

Which questions to answer?

Converting unstructured texts into structured data (with NLP & ML)

Sample Text from a Mammography Report

There is a **1.8cm round mass with a circumscribed margin** in the left breast in the anterior depth central to the nipple. Compared to previous films this mass is increase in size. There also is a **1.4cm oval mass with an obscured margin** in the left breast in the anterior depth of the inferior region. Compared to previous films **this mass** is increased in size.

Output of an ideal information extraction system

Imaging Observation 1: Mass

- **Size:** 1.8cm
- **Margin:** Circumscribed
- **Shape:** Round
- **Stability:** Increasing
- **Located In:** left breast in the anterior depth central to the nipple.

Imaging Observation 2: Mass

- **Size:** 1.4cm
- **Margin:** Obscured
- **Shape:** Oval
- **Stability:** Increasing
- **Located In:** left the left breast in the anterior depth of the inferior region

95%, 91%

PPV, Sensitivity for extracting imaging observations with their modifiers

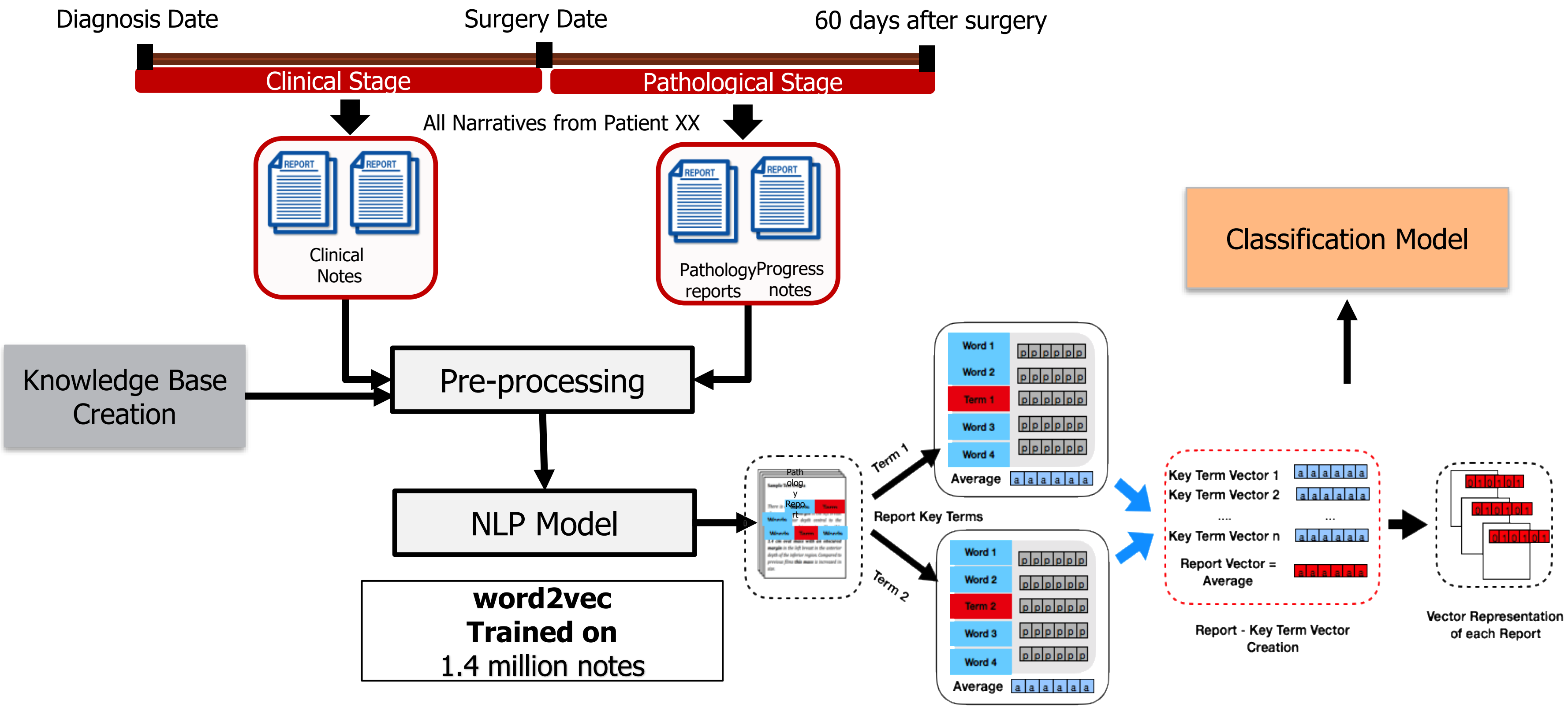
(1) Bozkurt, et al. Automatic abstraction of imaging observations with their characteristics from mammography reports. Journal of the American Medical Informatics Association 22.e1 (2015)

(2) Bozkurt, et al. Using automatically extracted information from mammography reports for decision-support. Journal of biomedical informatics 62 (2016)

Can we extract *missing* cancer stage data from clinical notes?

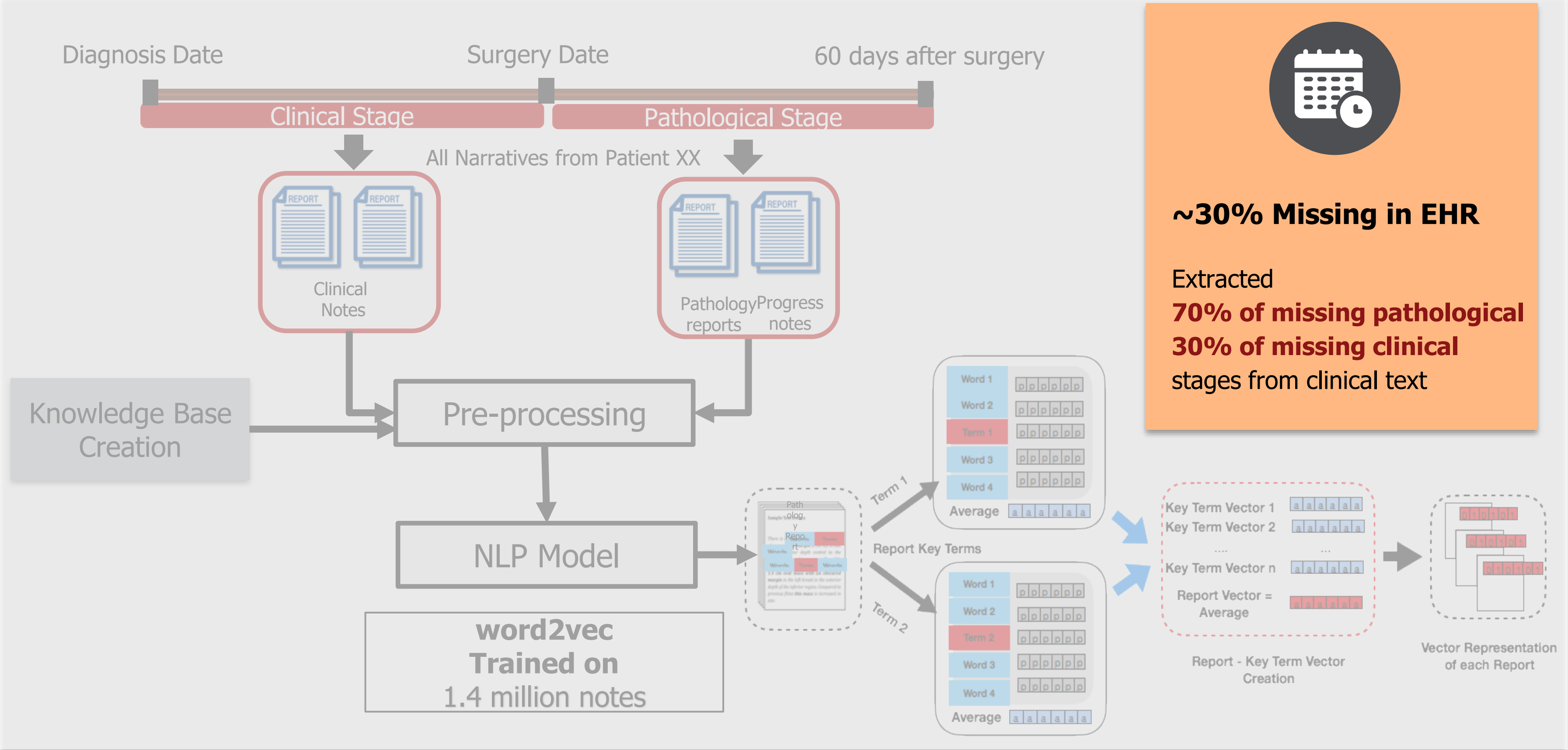
Stage information is *missing* from **10 to 50%** of patient records in cancer registries.

NLP pipeline



(3) Bozkurt et al. Expanding the Secondary Use of Prostate Cancer Real World Data: Automated Classifiers for Clinical and Pathological Stage. Frontiers in Digital Health (2022).

NLP pipeline results



~30% Missing in EHR

Extracted
70% of missing pathological
30% of missing clinical
stages from clinical text

(3) Bozkurt et al. Expanding the Secondary Use of Prostate Cancer Real World Data: Automated Classifiers for Clinical and Pathological Stage. Frontiers in Digital Health (2022).

Other examples: knowledge discovery from EHRs

	Examples
Detecting the Timeline of Recurrence	*Metastatic recurrence of breast cancer*
Phenotyping Severity of PCOs	*Urinary incontinence after chemo*
Detecting potential ambiguities	*Discrepancies between the laterality of an image and the reported impression.*

(4) Bozkurt et al. Phenotyping severity of patient-centered outcomes using clinical notes: A prostate cancer use case. Learning Health Systems (2020).
(5) Azad, A. D., Yilmaz, M., Bozkurt, S., Diverse patient trajectories during cytotoxic chemotherapy: Capturing longitudinal patient-reported outcomes. Cancer Medicine, 2020.
(6) Bozkurt S, et al. Automated detection of ambiguity in BI-RADS assessment categories in mammography reports. Cross-Border Challenges in Informatics with a Focus on Disease Surveillance and Utilizing Big Data

NLP and ML

unlock the power of EHRs at scale

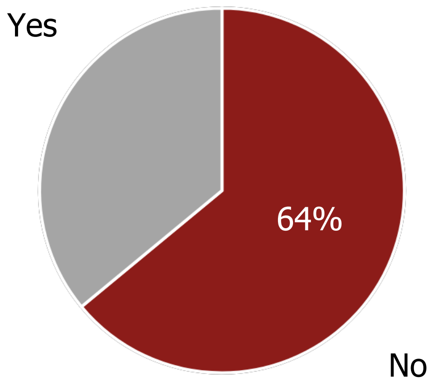
This involves making numerous careful decisions, rather than simply feeding all available data into a model and blindly accepting its output.

AI **Governance**

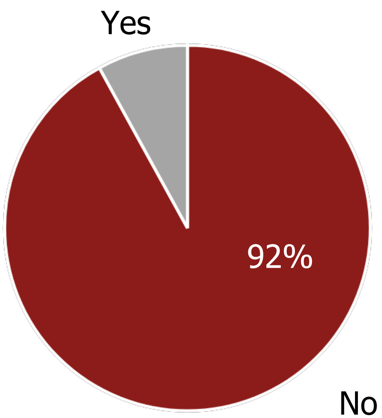
to ensure that these technologies are developed and deployed in a safe and responsible way

Reporting of demographic data and representativeness in ML models using EHR

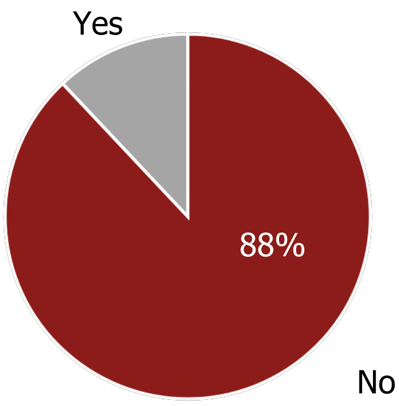
Race/Ethnicity



SES



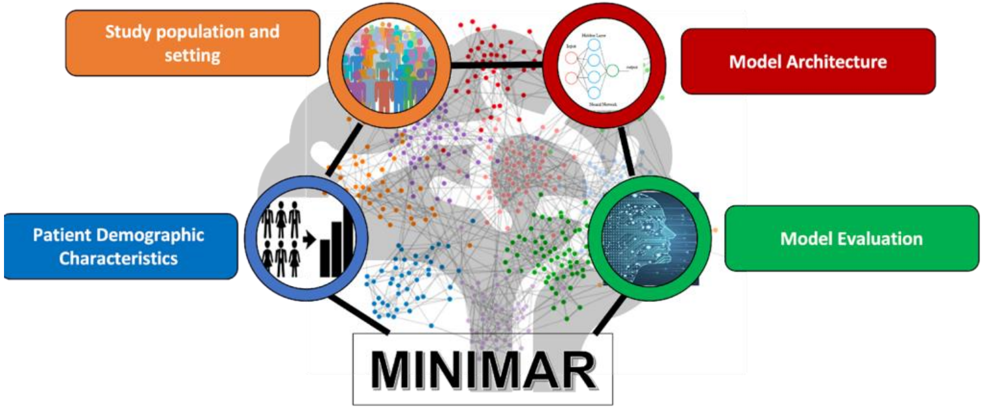
Prospective or External Validation



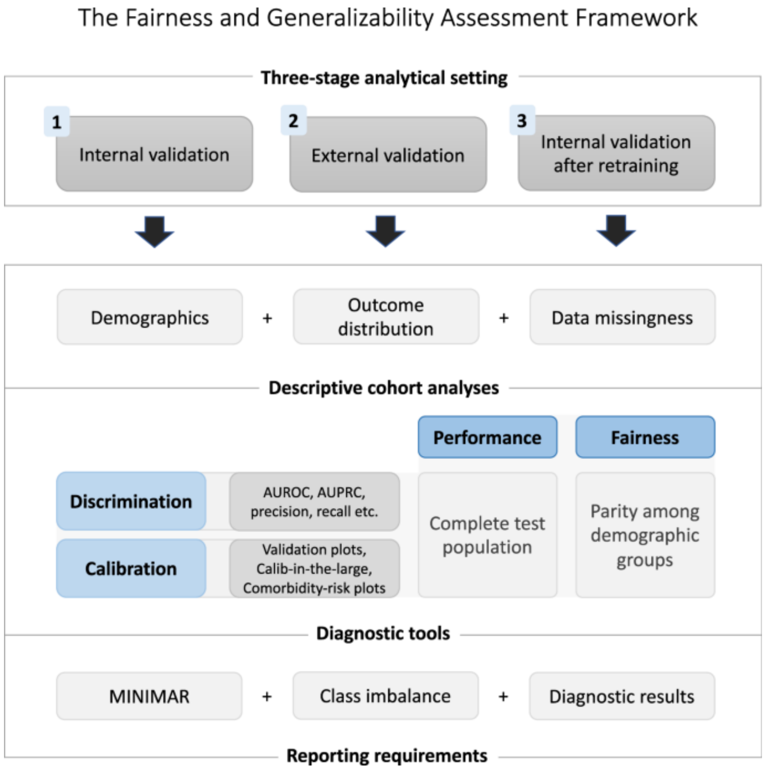
Bozkurt et al. "Reporting of demographic data and representativeness in machine learning models using electronic health records." Journal of the American Medical Informatics Association 27.12 (2020): 1878-1884.

AI best practices in healthcare

*As the number of models increases, it is becoming increasingly important to ensure that these models are **fair, unbiased, & generalizable**.*



MINimum **I**nformation for **M**edical **AI** **R**eporting

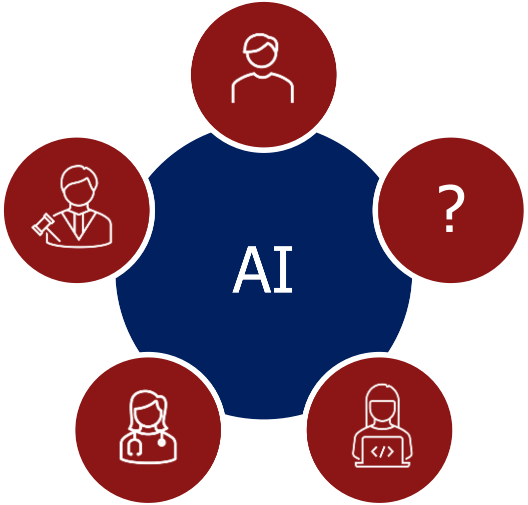


Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J. P., & Shah, N. H. (2020). MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. Journal of the American Medical Informatics Association, 27(12), 2011-2015.

AI Governance in healthcare

Lifecycle and Key Dimensions of an AI System

A socio-technical team



Thank you!

Contact

selenb@stanford.edu

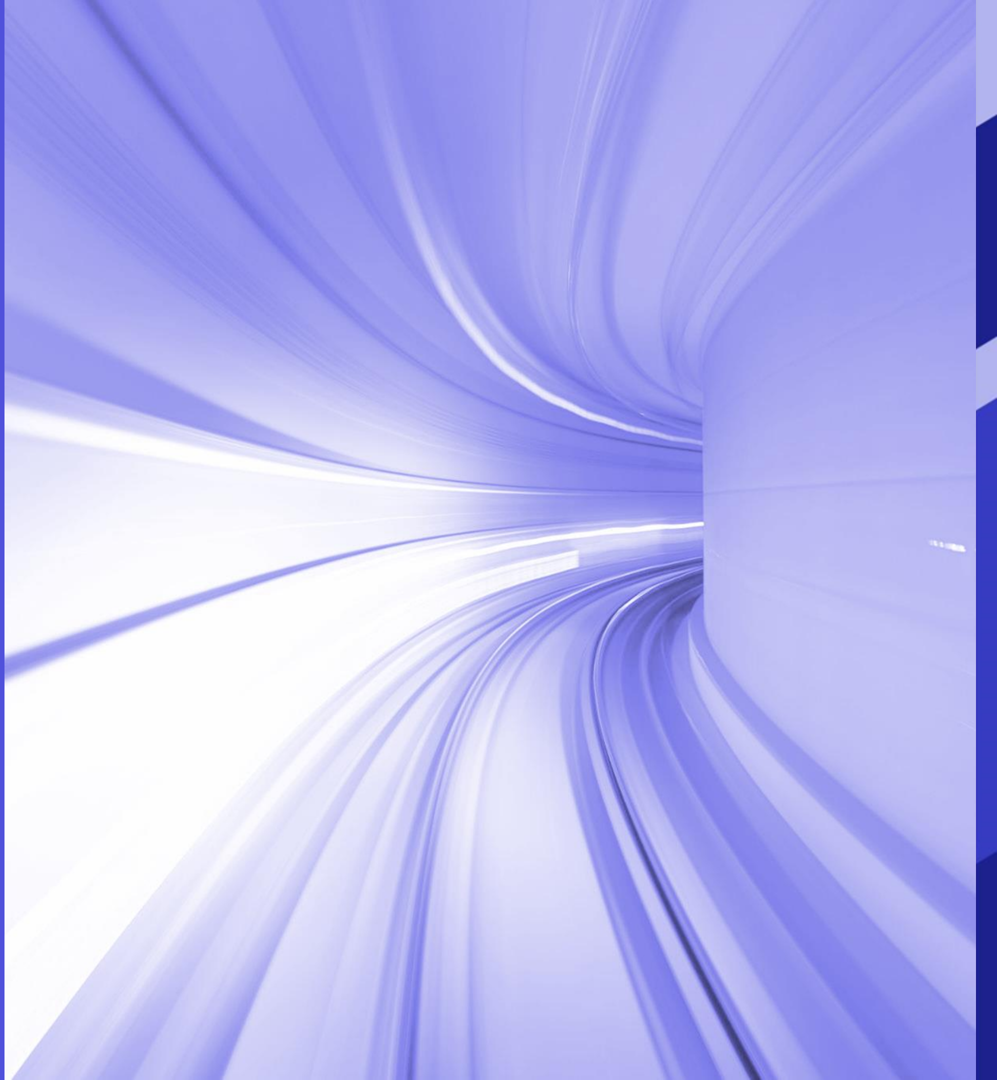
selenbw.github.io

I'm hiring!
Looking for interns, grad students and postdocs

ML-enabled data extraction, faster

Katherine Tan
Flatiron Health

@statistikat
wlktan



flatiron

Disclaimers

Katherine Tan is an employee of Flatiron Health, an independent subsidiary of Roche Group.

She holds stock ownership in Roche.

Machine Learning can enhance RWE for HEOR

Cohort selection, identifying samples with greater specificity with respect to inclusion criteria

Identification of of independent **predictors and covariates of health outcomes**

And much more…

ELSEVIER

Value in Health

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

ISPOR Report

Machine Learning Methods in Health Economics and Outcomes Research—The PALISADE Checklist: A Good Practices Report of an ISPOR Task Force

William V. Padula, PhD, Noemi Kreif, PhD, David J. Vanness, PhD, Blythe Adamson, PhD, Juan-David Rueda, MD, PhD, Federico Felizzi, PhD, MBA, Pall Jonsson, PhD, Maarten J. IJzerman, PhD, Atul Butte, MD, PhD, William Crown, PhD

ABSTRACT

Advances in machine learning (ML) and artificial intelligence offer tremendous potential benefits to patients. Predictive analytics using ML are already widely used in healthcare operations and care delivery, but how can ML be used for health economics and outcomes research (HEOR)? To answer this question, ISPOR established an emerging good practices task force for the application of ML in HEOR.

The task force identified 5 methodological areas where ML could enhance HEOR: (1) cohort selection, identifying samples with greater specificity with respect to inclusion criteria; (2) identification of independent predictors and covariates of health outcomes; (3) predictive analytics of health outcomes, including those that are high cost or life threatening; (4) causal inference through methods, such as targeted maximum likelihood estimation or double-debiased estimation—helping to produce reliable evidence more quickly; and (5) application of ML to the development of economic models to reduce structural, parameter, and sampling uncertainty in cost-effectiveness analysis.

Overall, ML facilitates HEOR through the meaningful and efficient analysis of big data. Nevertheless, a lack of transparency on how ML methods deliver solutions to feature selection and predictive analytics, especially in unsupervised circumstances, increases risk to providers and other decision makers in using ML results.

To examine whether ML offers a useful and transparent solution to healthcare analytics, the task force developed the PALISADE Checklist. It is a guide for balancing the many potential applications of ML with the need for transparency in methods development and findings.

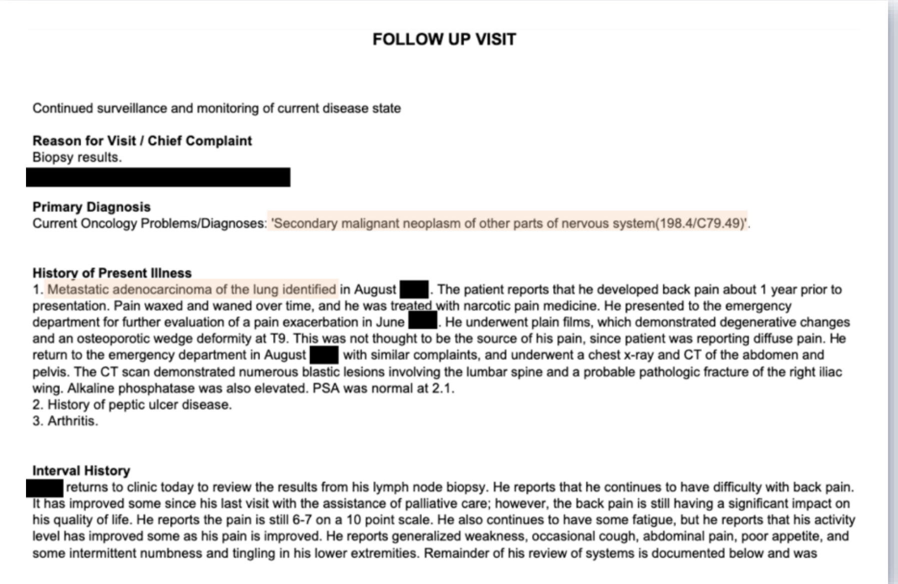
Keywords: artificial intelligence, machine learning.

Padula W, et al. Machine Learning Methods in Health Economics and Outcomes Research—The PALISADE Checklist: A Good Practices Report of an ISPOR Task Force. ISPOR Report. https://doi.org/10.1016/j.jval.2022.03.022

Challenge: Critical data elements come from unstructured data

Several data elements critical for outcomes research are stored in unstructured data sources.

Abstracting this information is a **costly** and **resource intensive** task.


FOLLOW UP VISIT

Continued surveillance and monitoring of current disease state

Reason for Visit / Chief Complaint
Biopsy results.
[redacted]

Primary Diagnosis
Current Oncology Problems/Diagnoses: 'Secondary malignant neoplasm of other parts of nervous system(198.4/C79.49)'.

History of Present Illness
1. Metastatic adenocarcinoma of the lung identified in August [redacted]. The patient reports that he developed back pain about 1 year prior to presentation. Pain waxed and waned over time, and he was treated with narcotic pain medicine. He presented to the emergency department for further evaluation of a pain exacerbation in June [redacted]. He underwent plain films, which demonstrated degenerative changes and an osteoporotic wedge deformity at T9. This was not thought to be the source of his pain, since patient was reporting diffuse pain. He return to the emergency department in August [redacted] with similar complaints, and underwent a chest x-ray and CT of the abdomen and pelvis. The CT scan demonstrated numerous blastic lesions involving the lumbar spine and a probable pathologic fracture of the right iliac wing. Alkaline phosphatase was also elevated. PSA was normal at 2.1.
2. History of peptic ulcer disease.
3. Arthritis.

Interval History
[redacted] returns to clinic today to review the results from his lymph node biopsy. He reports that he continues to have difficulty with back pain. It has improved some since his last visit with the assistance of palliative care; however, the back pain is still having a significant impact on his quality of life. He reports the pain is still 6-7 on a 10 point scale. He also continues to have some fatigue, but he reports that his activity level has improved some as his pain is improved. He reports generalized weakness, occasional cough, abdominal pain, poor appetite, and some intermittent numbness and tingling in his lower extremities. Remainder of his review of systems is documented below and was

Tradeoffs building EHR data solutions using traditional approach of manual chart review

Scale and Size

Limited cohort size, particularly for nuanced cohorts

Recency and Speed

Lag between research question and availability of temporally recent data

Clinical Depth

Relevant clinical details that are difficult to abstract at scale are missing

Deep learning models mimic human abstraction

- **Scalable, automated** extraction of clinical concepts **explicitly documented** in the patient chart
- **It is NOT:** prediction or inference of a clinical value based on other patient characteristics; generative modeling (ie, different from chat GPT)

Deep Learning Model Example	Language in Source EHR as Illustrative Snippet (Model Input)	Extracted Variables (Model Outputs)
Biomarkers	"Mr. Smith received NGS test results on 2/20/2021 for EGFR, ALK, and ROS1 and was found to have an ALK rearrangement."	BiomarkerName, BiomarkerStatus, ResultReturnedDate
PD-L1	"Mr. Smith was diagnosed with adenocarcinoma of the lung, PD-L1 <1% on 2/20/2021."	BiomarkerName, PercentStaining, ResultReturnedDate
KRAS	"Mr. Smith tested positive for a KRAS G12C mutation on 1/15/2019."	MutationG12CDetail

Adamson B, et al. Methods for machine learning extraction of RWD variables from electronic health records. medRxiv 2023.03.02.23286522

Adding KRAS G12C mutation details to a lung cancer dataset using ML

In addition to **biomarker testing status**, knowing mutation details became important for researchers as SOC evolved and targeted therapies are approved

Patient	Biomarker	Result date	G12C mutation?	Sample type	Tissue Collection site	Test type	…
A	KRAS	2015-03-19	?	Tissue	Primary	NGS	
B	KRAS	2017-01-23	?	Tissue	Metastatic	NGS	

ML-extracted biomarker mutation detail

EXAMPLE 1:

KRAS G12C

Name: John Smith
marker value result egfr (mutation) wild-type no mutation **[kras] [gl2c]** mutation <u>positive</u> ' table 1: summary of genes, values and results

EXAMPLE 2:

Other point mutation

Name: Jane Doe
marker value result rrmi 0.37 low expression pdl1 high expression (60%) **[BRAF] [V600E]** <u>positive</u>. ' table 1: summary of genes, values and results

Generalized Biomarker Models

EXAMPLE 1:

KRAS G12C

Name: John Smith
marker value result egfr (mutation) wild-type no mutation **[kras] [gl2c]** mutation <u>positive</u> ' table 1: summary of genes, values and results

↓

Name: John Smith
marker value result egfr (mutation) wild-type no mutation **[BIOMARKER] [MUTATION]** mutation <u>positive</u> ' table 1: summary of genes, values and results

EXAMPLE 2:

Other point mutation

Name: Jane Doe
marker value result rrmi 0.37 low expression pdl1 high expression (60%) **[BRAF] [V600E]** <u>positive</u>. ' table 1: summary of genes, values and results

↓

Name: Jane Doe
marker value result rrmi 0.37 low expression pdl1 high expression (60%) **[BIOMARKER] [MUTATION]** <u>positive</u>. ' table 1: summary of genes, values and results

ML-extraction enabled velocity of obtaining insights at scale and depth

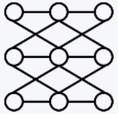
ABSTRACTION

13,000

tasks

(3,700 hours)

ML-EXTRACTION



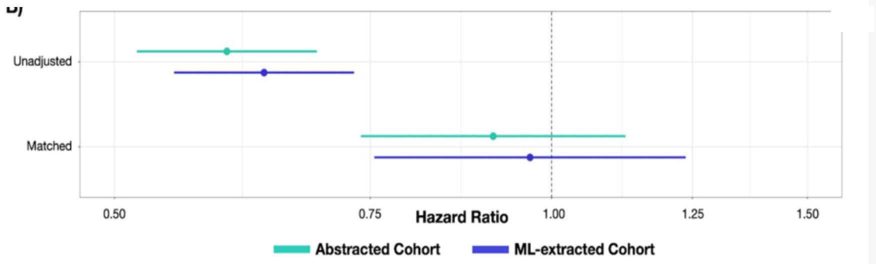
1,400

tasks

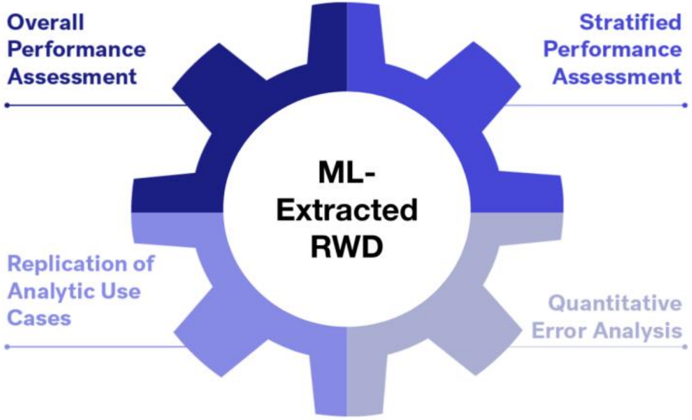
(10% of the abstraction load)

ML-extracted data can generate similar results and conclusions as abstracted data

Results from replication of natural history study of biomarker associated survival



Benedum C, et al. Replication of Real-World Evidence in Oncology Using Electronic Health Record Data Extracted by Machine Learning. *Cancers.* 2023;15(6). doi:10.3390/cancers15061853



Estevez M, et al. Considerations for the use of machine learning extracted real-world data to support evidence generation: A research-centric evaluation framework. *Cancers*. 2022;14(13). doi: 10.3390/cancers14133063

Thank you

Additional Collaborators: Erin Fidyk, Blythe Adamson, Chaya Wurman, Melissa Estevez, Sheila Nemeth, Catherine Au-Yeung (Design)

Katherine Tan

Senior Quantitative Scientist

Machine Learning & Data Capabilities

Flatiron Health

@statistikat

wlktan

flatiron

Why Human-Algorithm Collaborations will Transform Care Delivery

Ravi B. Parikh, MD, MPP

Ravi.parikh@pennmedicine.upenn.edu

@ravi_b_parikh

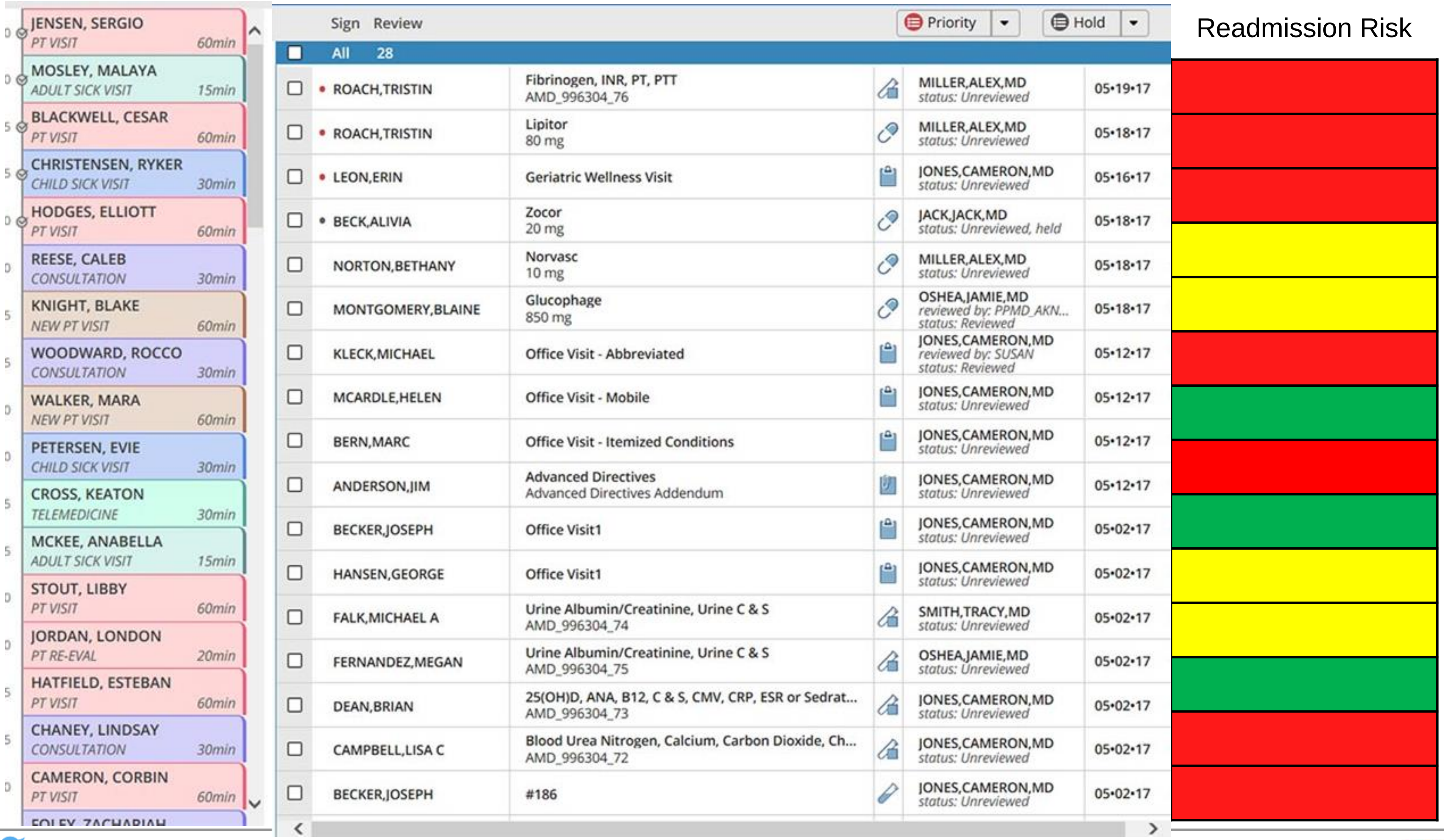
HACLAB

Penn Medicine

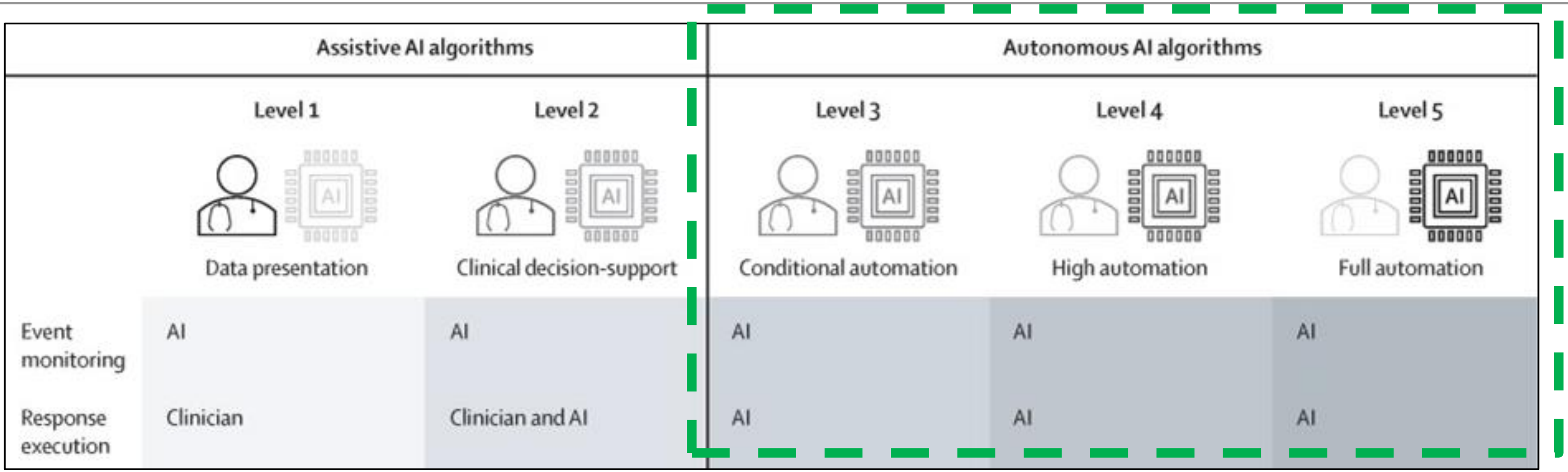
Disclosures

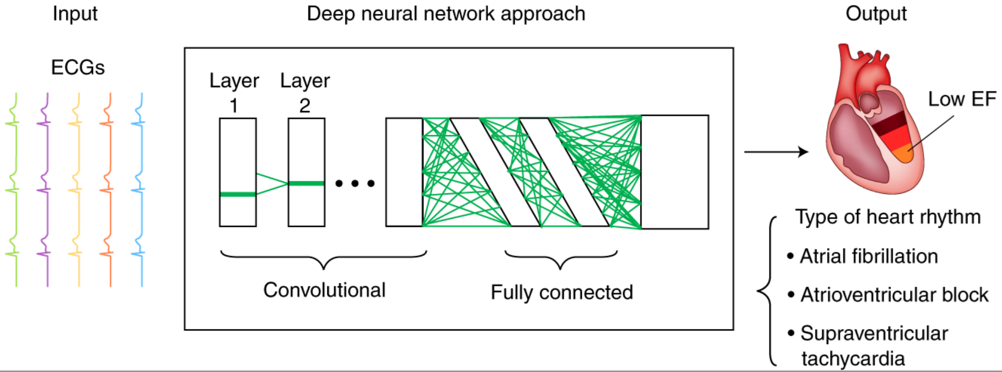
- **<u>Research Funding</u>: NIH, Department of Defense, Veterans Health Administration, National Palliative Care Research Center, Humana, Prostate Cancer Foundation, NCCN Foundation, Conquer Cancer Foundation, Emerson Collective**
- **<u>Consultant:</u> Thyme Care, Onc.ai, Biofourmis, Humana, Cancer Study Group, GNS Healthcare, Nanology**
- **<u>Columnist:</u> Medscape, Flatiron**
- **<u>Leadership</u>: Coalition to Transform Advanced Care, National Quality Forum**

My first experience with AI



We pretend like all AI is autonomous…





HACLAB

…when instead most current AI is assistive

	Assistive AI algorithms		Autonomous AI algorithms		
	Level 1	Level 2	Level 3	Level 4	Level 5
	Data presentation	Clinical decision-support	Conditional automation	High automation	Full automation
Event monitoring	AI	AI	AI	AI	AI
Response execution	Clinician	Clinician and AI	AI	AI	AI

Human-algorithm collaborations require more than data science

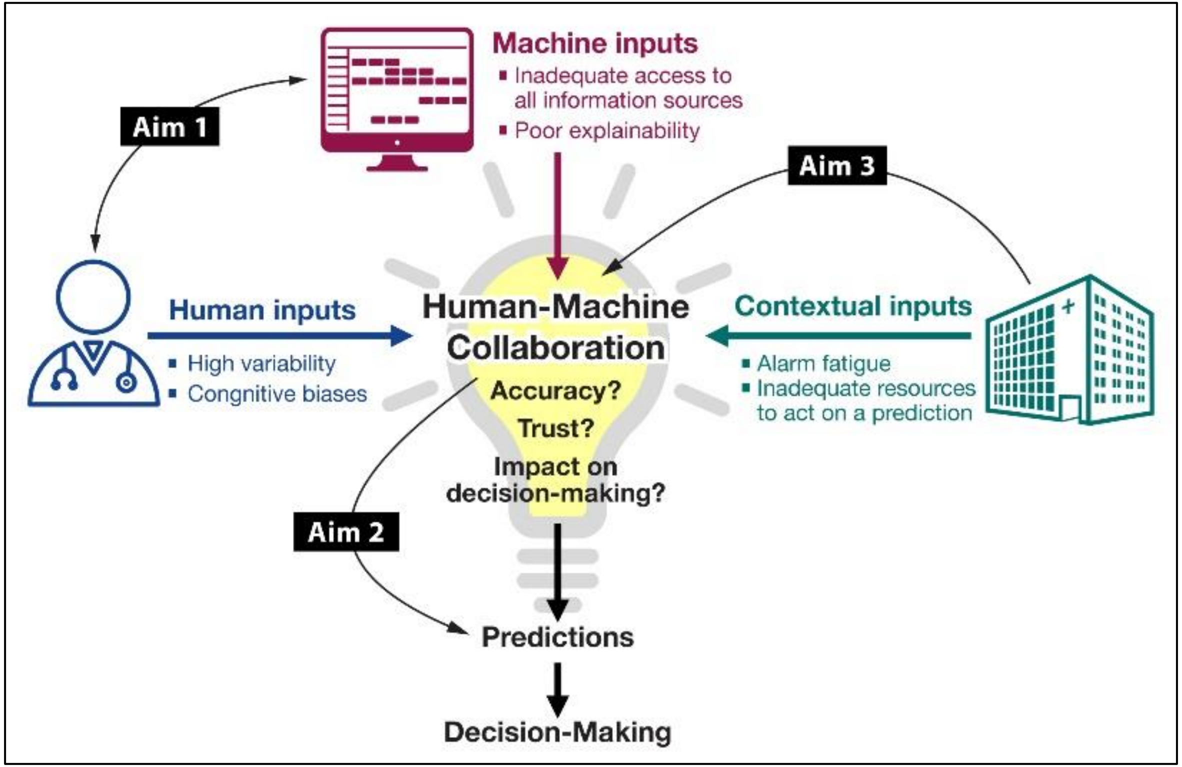
Data Science

Computer Science

Behavioral Science

Clinicians

Mixed-Methods



Human Factors

Implementation Science

Operational leaders

HACLAB

A use case: Serious Illness Communication

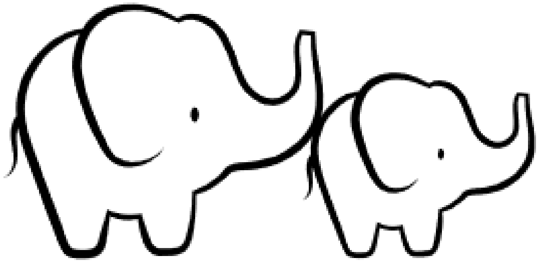
- Early communication is key to reducing unwarranted end-of-life care in oncology
- Identifying appropriate patients is key, but oncologists do badly at prognosis

Why do existing solutions fail in real-world practice?

Identifying the right patients is hard

Machine Learning

Changing behavior is hard

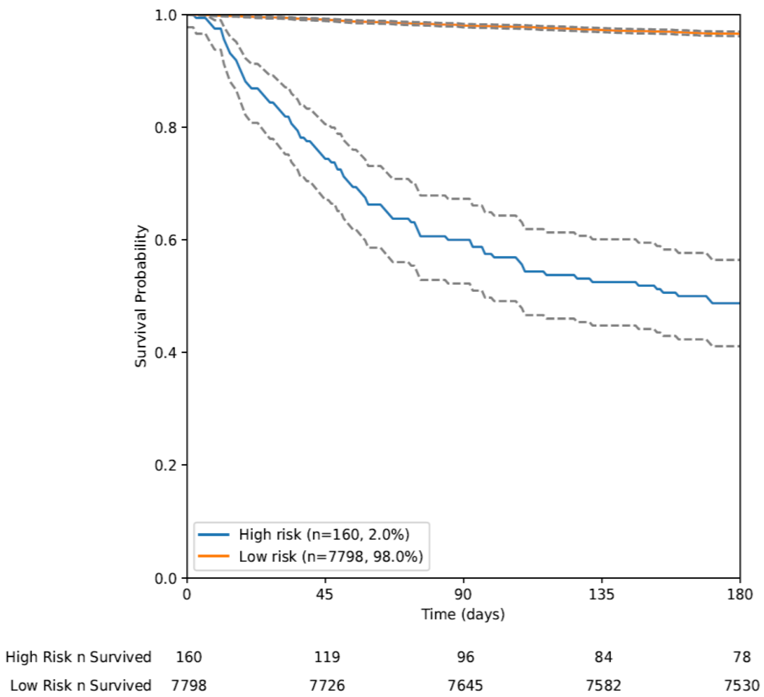


Behavioral Economics

HACLAB

Christakis and Lamont, BMJ, 2003; Bestvina and Polite, JOP, 2017

Algorithm development and validation



Clinician perspectives on machine learning prognostic algorithms in the routine care of patients with cancer: a qualitative study

Original Investigation | Oncology

Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer

JAMA Oncology | Original Investigation

Validation of a Machine Learning Algorithm to Predict 180-Day Mortality for Outpatients With Cancer

Variables	Examples	Features
Demographics	Age, Gender	
Comorbidities	33 Elixhauser comorbidities	• Total count • Recent*
Cancer-specific	Stage, tumor markers	• Total count • First/last value • Min/Max • Proportion ordered STAT
Laboratories	CMP, CBC, LDH	
Recent utilization	Outpatient visit number	

HACLAB

Incorporating behavioral economics with machine learning

Use Case: Predicting mortality to prompt more serious illness communication

Next week's high-risk patients for Serious Illness Conversations

Thursday, July 18, 2019 at 8:05 AM

Show Details

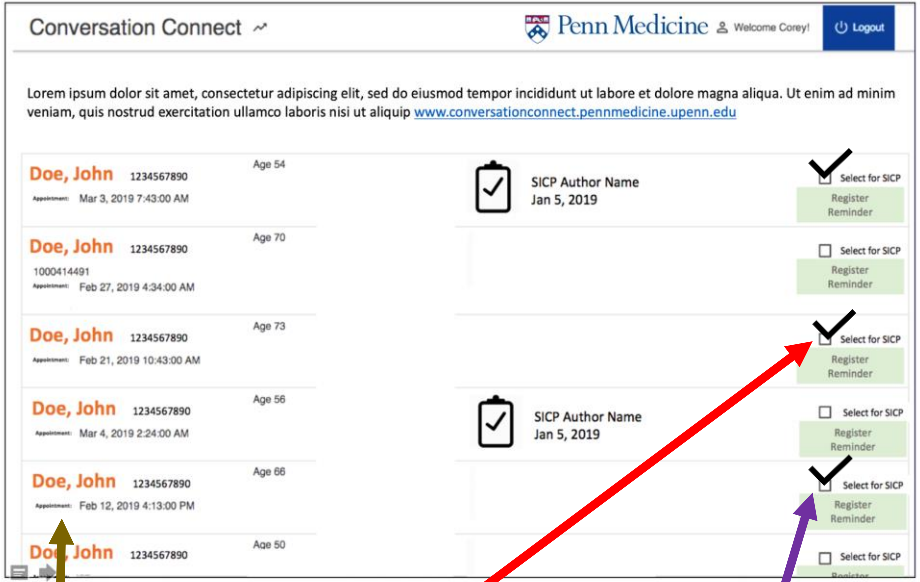
Dear :

The ACC is working to help oncologists have earlier Serious Illness Conversations with patients. In the past four weeks, you have documented **2 conversations.**

18 oncology clinicians have documented more conversations than you during that time.

We have identified patients scheduled to see you next week who may benefit from a Serious Illness Conversation. Click here to view your list (you must be connected to the UPHS network).

Sincerely,



Peer Comparison

Performance Report

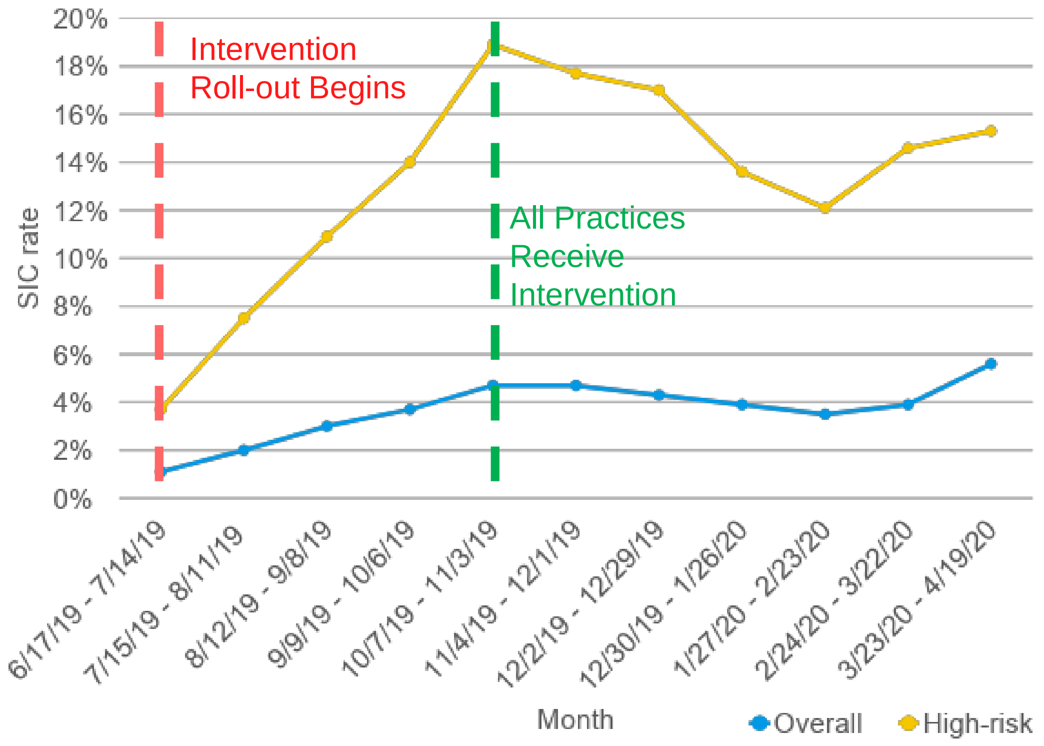
ML high-risk list

Pre-commitment

Default Text Message

HACLAB

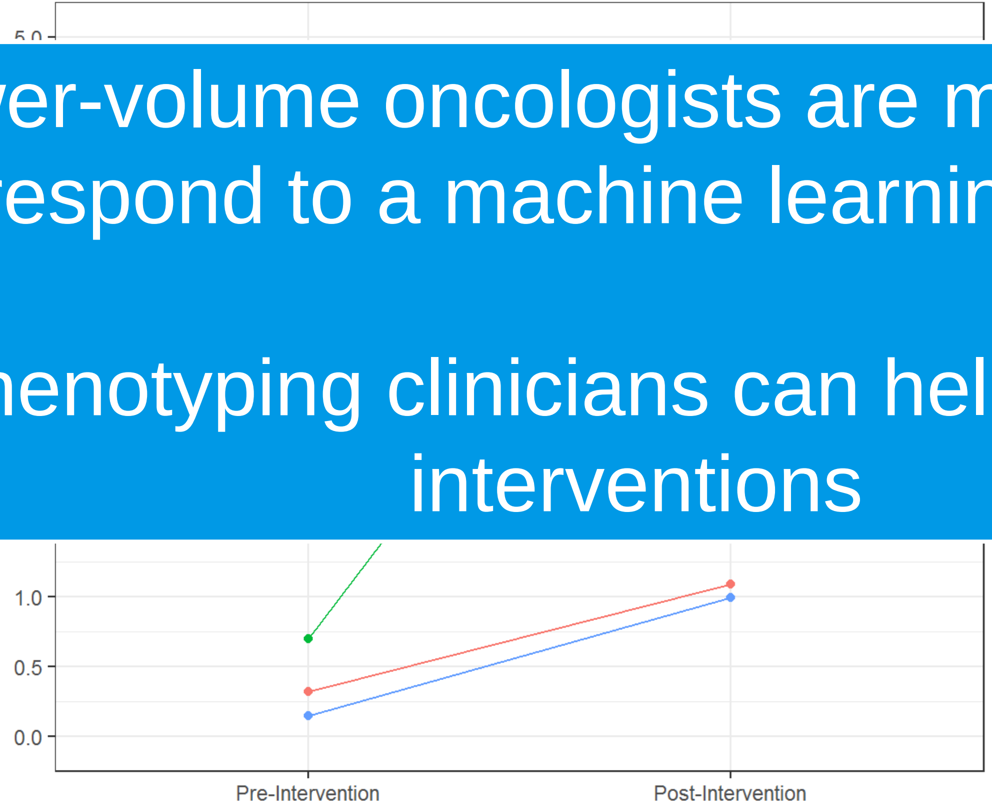
Conversation Connect Impact



	Control	Intervention
Chemo last 14 days*	**10.4%**	**7.5%**
Hospice before death	59.6%	60.6%
ICU last 30 days	16.9%	15.7%
Spending last 6 months of life*	$65,971.09	$78,256.24

HACLAB

Humans don’t respond to AI in the same way



Lower-volume oncologists are more likely to respond to a machine learning nudge

Phenotyping clinicians can help refine AI interventions

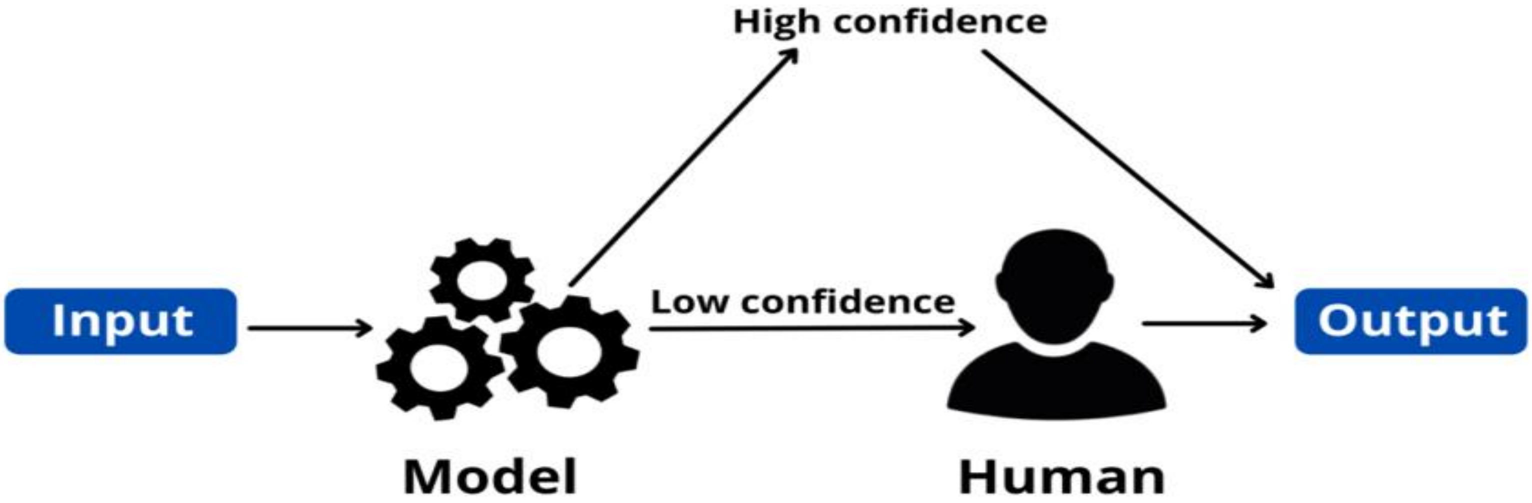
Li et al, PLOS One, 2022

HACLAB

Human-centered AI can mitigate disparities

	Pre-intervention, %	Post-intervention, %	Absolute Percentage-point Difference
Non-Hispanic White	3.9 (58/1494)	14.2 (201/1417)	10.3
Non-Hispanic Black	3.6 (17/467)	16.9 (69/408)	13.3
Other*	1.2 (2/164)	19.5 (34/408)	18.3

Human-in-the-loop models are a promising strategy



Lessons Learned

Machine learning predictions can improve care <u>when</u>

- Clinicians' perspectives are solicited prior to algorithm development
- The algorithm is "vetted" prior to implementation
- They are well-integrated into clinical workflows
- They are paired with behavioral nudges rather than simply displayed on a computer screen

Thank you!

Ravi.parikh@pennmedicine.upenn.edu
@ravi_b_parikh
www.ravibparikh.com

Dana-Farber Cancer Institute
Christopher Manz

Abramson Cancer Center
Lynn Schuchter
Larry Shulman
Jennifer Braun
Pallavi Kumar

Penn Data Science
Michael Draugelis
Corey Chivers
Thaibinh Luong
Susan Harkness Regli

Penn Nudge Unit
Mitesh Patel
Sujatha Changolkar
Charles Rareshide
Chalanda Evans

Penn Center for Cancer Care Innovation
Justin Bekelman
Sam Takvorian
Katherine Rendle

Mixed-Methods Research Lab
Judy Shea
Maria Nelson

Leonard Davis Institute/Dept of Medical Ethics and Health Policy
Manqing Liu
Yichen Zhang
Will Ferrell
Amol Navathe
Ezekiel Emanuel
Kevin Volpp

UC Berkeley
Ziad Obermeyer

NIH NATIONAL CANCER INSTITUTE

PENN CENTER FOR PRECISION MEDICINE

NATIONAL Palliative Care RESEARCH CENTER

ITMAT Institute for Translational Medicine and Therapeutics

HACLAB