With great power comes great responsibility:

Machine learning in clinical research

E. Hope Weissler, MD, MHS & Erich Huang, MD, PhD

Disclosures

Weissler: None

Huang:
Founder - kelaHealth, Clinetic, Stratus Medicine
Chief Science and Innovation Officer at Onduo (a Verily company)
Advisor – Optimizely

Contents

Past

What is holding clinical research back from reaching full potential?

Contents

Past

What is holding clinical research back from reaching full potential?

Present

How might machine learning address those issues?

What are the barriers to ML implementation in CR?

Contents

Past

What is holding clinical research back from reaching full potential?

Present

How might machine learning address those issues?

What are the barriers to ML implementation in CR?

Future

How can we overcome these barriers?

Contents

DCRI **THINK TANKS**
FROM INSIGHT TO ACTION

Leveraging Artificial Intelligence and Machine Learning Methods and Approaches to Transform Clinical Trial Design, Planning, and Execution

~50 attendees including:

- clinical researchers
- machine learning experts
- biopharmaceutical industry
- technology companies
- patient advocacy groups
- FDA

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Trials

COMMENTARY Open Access

Check for updates

The role of machine learning in clinical research: transforming the future of evidence generation

E. Hope Weissler[1*], Tristan Naumann[2], Tomas Andersson[3], Rajesh Ranganath[4], Olivier Elemento[5], Yuan Luo[6], Daniel F. Freitag[7], James Benoit[8], Michael C. Hughes[9], Faisal Khan[3], Paul Slater[10], Khader Shameer[3], Matthew Roe[11], Emmette Hutchison[3], Scott H. Kollins[1], Uli Broedl[12], Zhaoling Meng[13], Jennifer L. Wong[14], Lesley Curtis[1], Erich Huang[1,15] and Marzyeh Ghassemi[16,17,18,19]

Ref 1.

Present: What can we do with ML in CR?

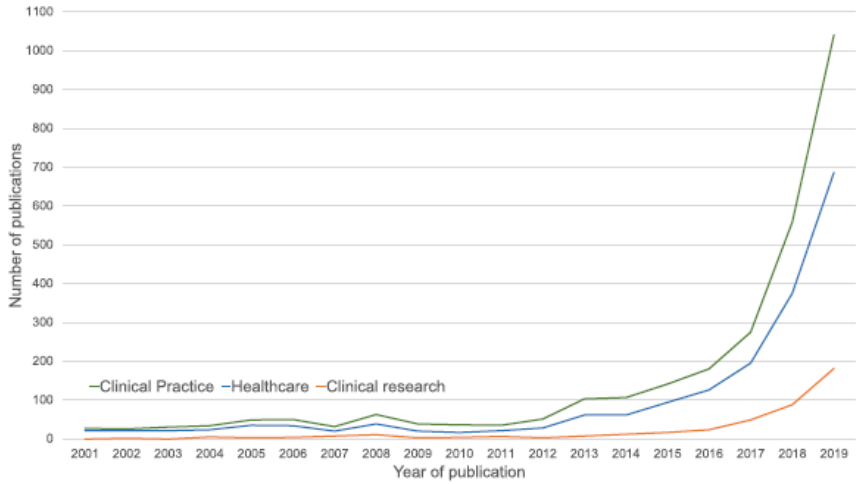


Fig. 1 The number of clinical practice–related publications was determined by searching "("machine learning" or "artificial intelligence") and ("healthcare")." The number of healthcare-related publications was determined by searching "("machine learning" or "artificial intelligence") and ("healthcare")", and the number of clinical research–related publications was determined by searching "("machine learning" or "artificial intelligence") and ("clinical research")."

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

BRIEF COMMUNICATION OPEN

The National Institutes of Health funding for clinical research applying machine learning techniques in 2017

Amarnath R. Annapureddy[1,2], Suveen Angraal[1,3], Cesar Caraballo[1], Alyssa Grimshaw[4], Chenxi Huang[1], Bobak J. Mortazavi[5] and Harlan M. Krumholz[1,2,6*]

Ref. 2

Contents

Prior reviews of the topic

Circulation: Cardiovascular Quality and Outcomes

NOVEL STATISTICAL METHODS

Recommendations for Reporting Machine Learning Analyses in Clinical Research

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

ELSEVIER

A survey of machine learning applications in HIV clinical research and care

Kuteesa R. Bisaso[a,b,c,*], Godwin T. Anguzu[a], Susan A. Karungi[b], Agnes Kiragga[a], Barbara Castelnuovo[a]

REVIEW

Korean J Intern Med 2019;34:708-722
https://doi.org/10.3904/kjim.2018.349

KJIM

Application of machine learning in rheumatic disease research

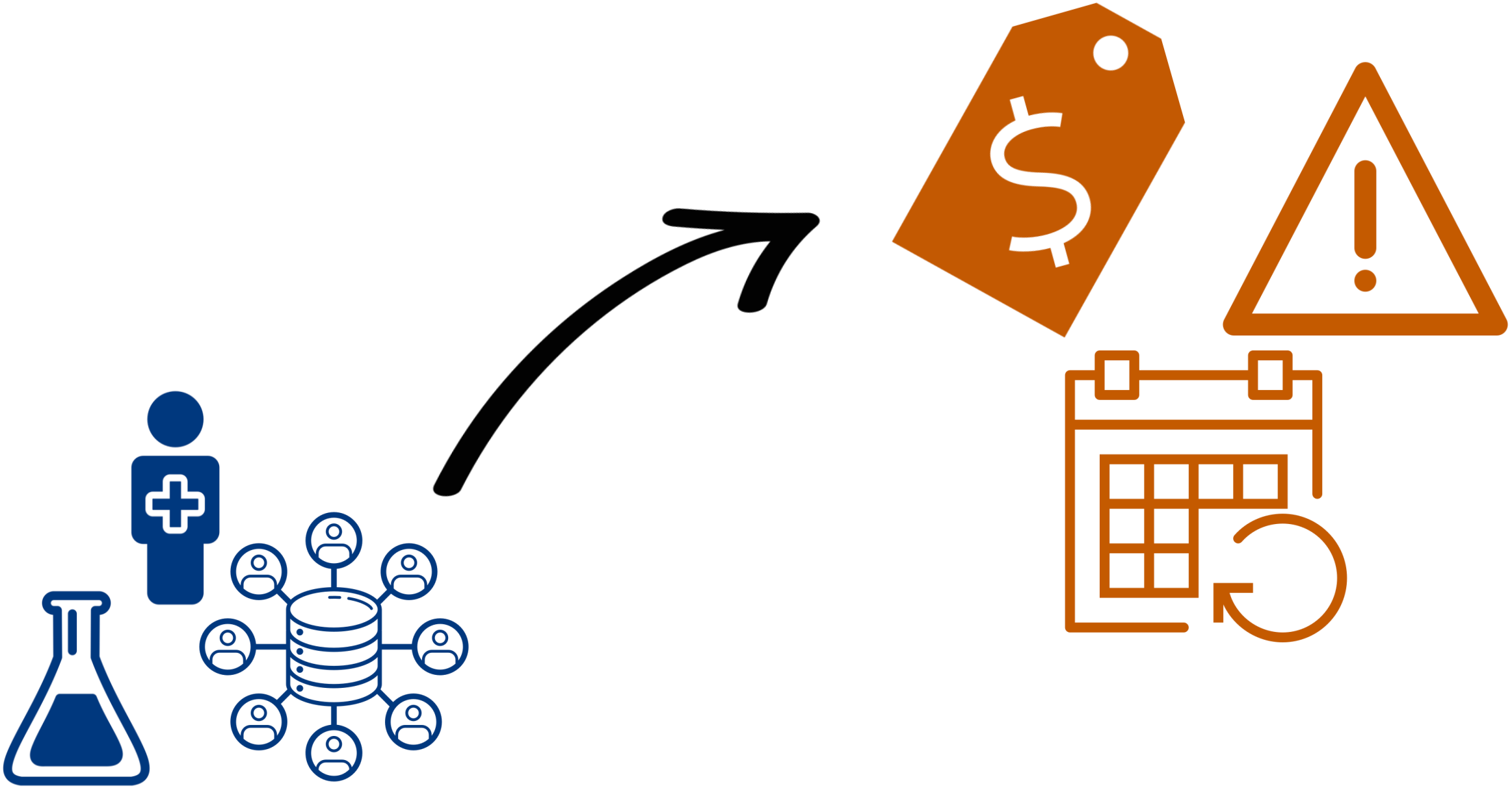
Ki-Jo Kim[1] and Ilias Tagkopoulos[2,3]

Ref. 3-5

Pre-clinical work | **Study planning** | **Study conduct: Recruitment, retention, data collection** | **Data analysis**

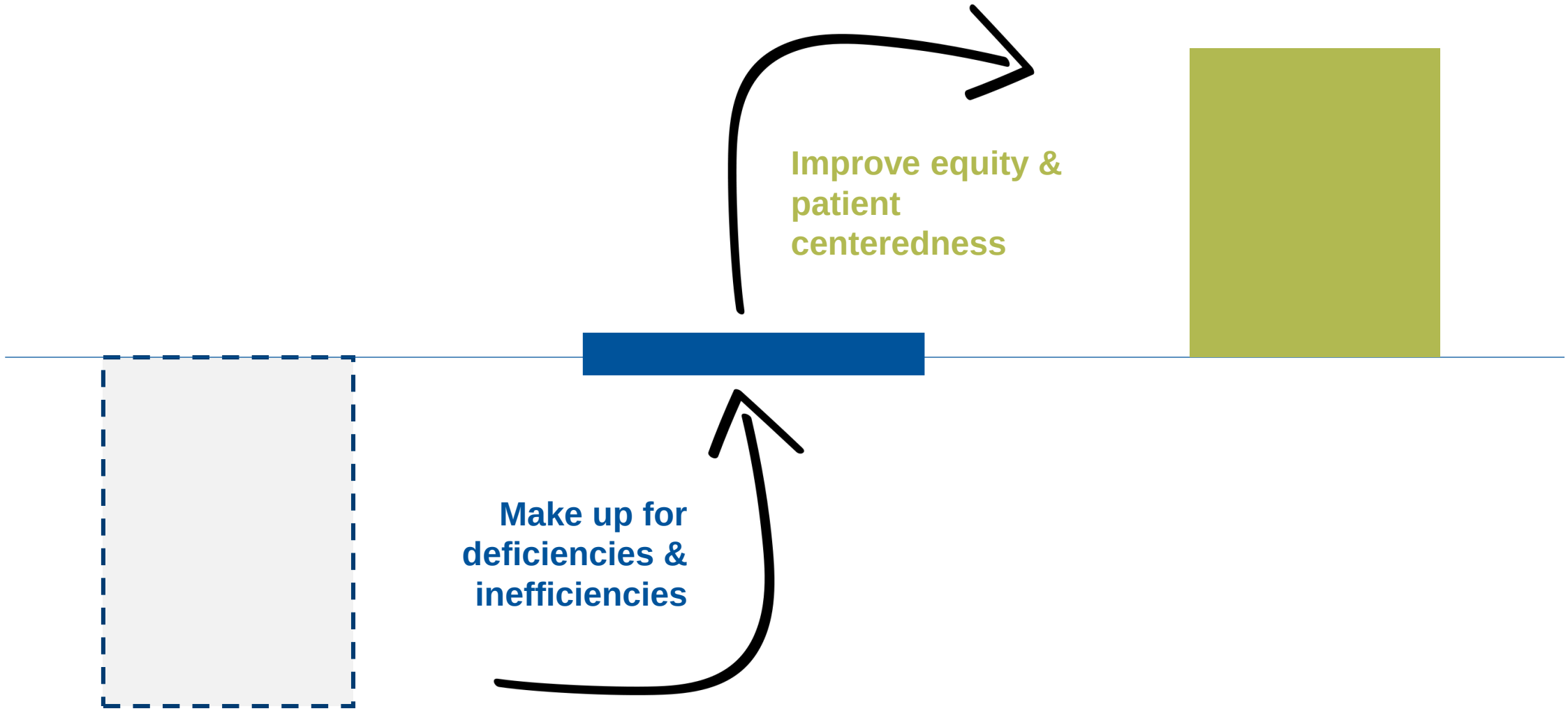
What we'll talk about today

Past: Opportunities for clinical research improvement

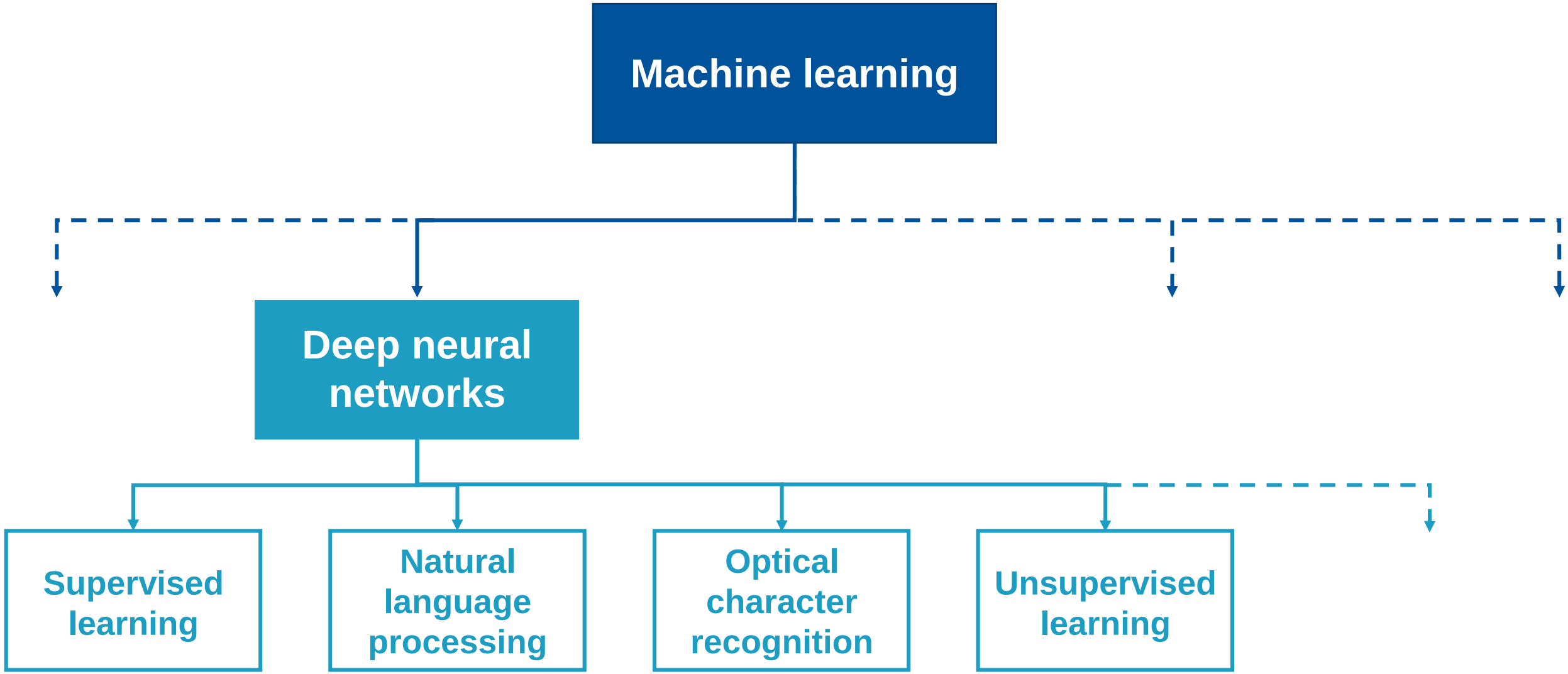


Ref. 6, Wong

Past: Opportunities for clinical research improvement



Present: What can we do with ML in CR?



Present: What can we do with ML in CR?

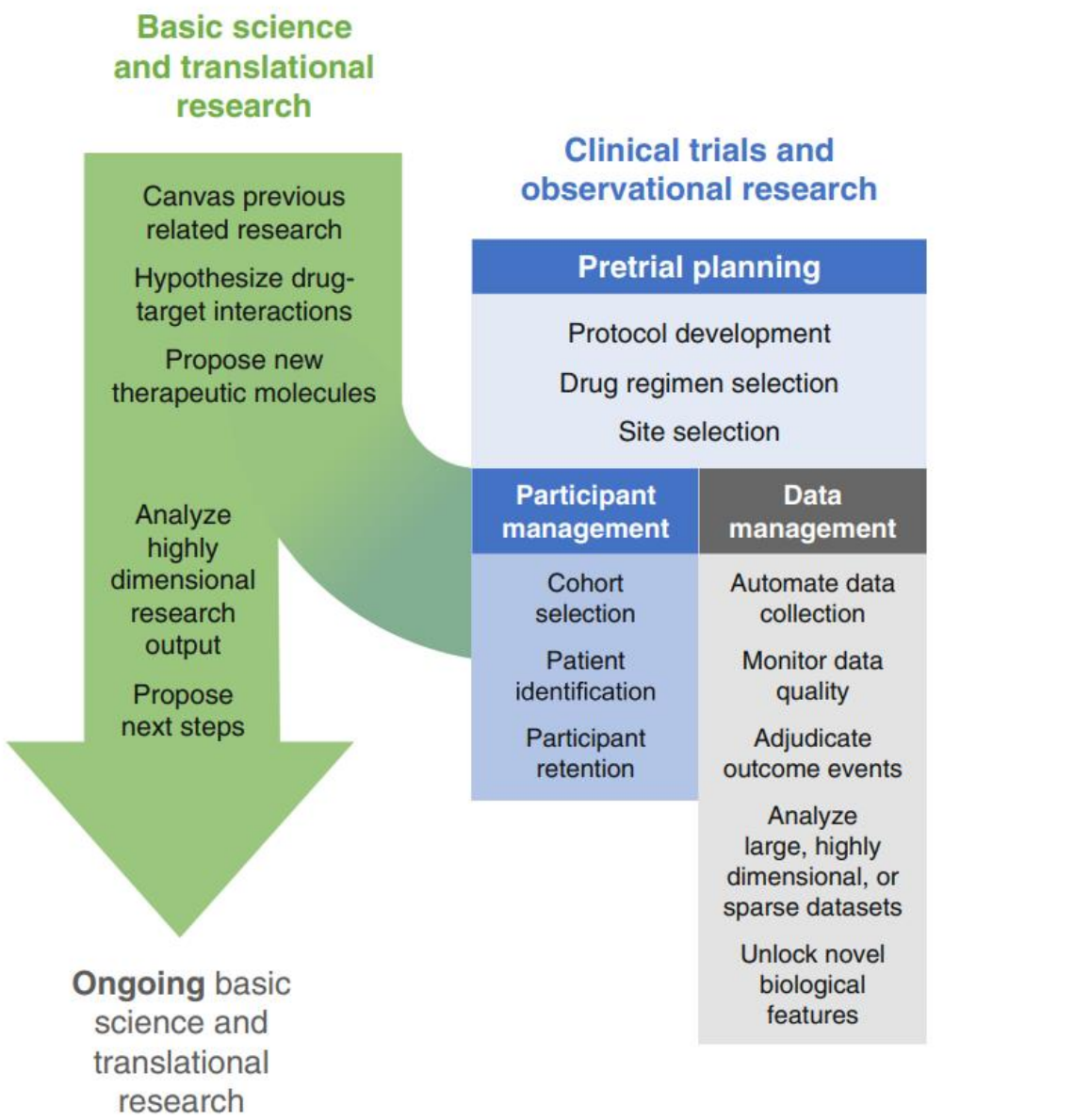


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

Expensive
Time consuming
Prone to false starts

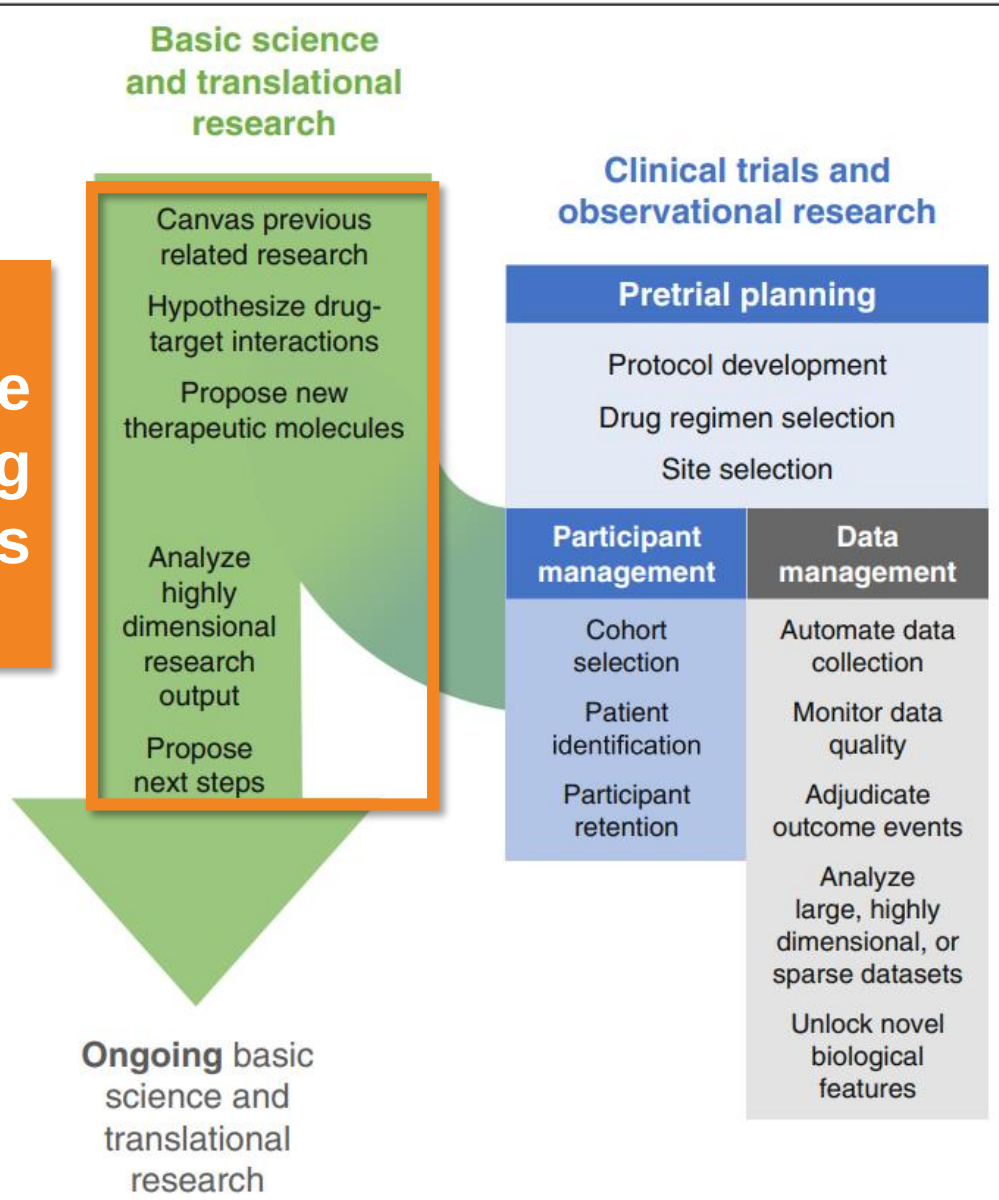
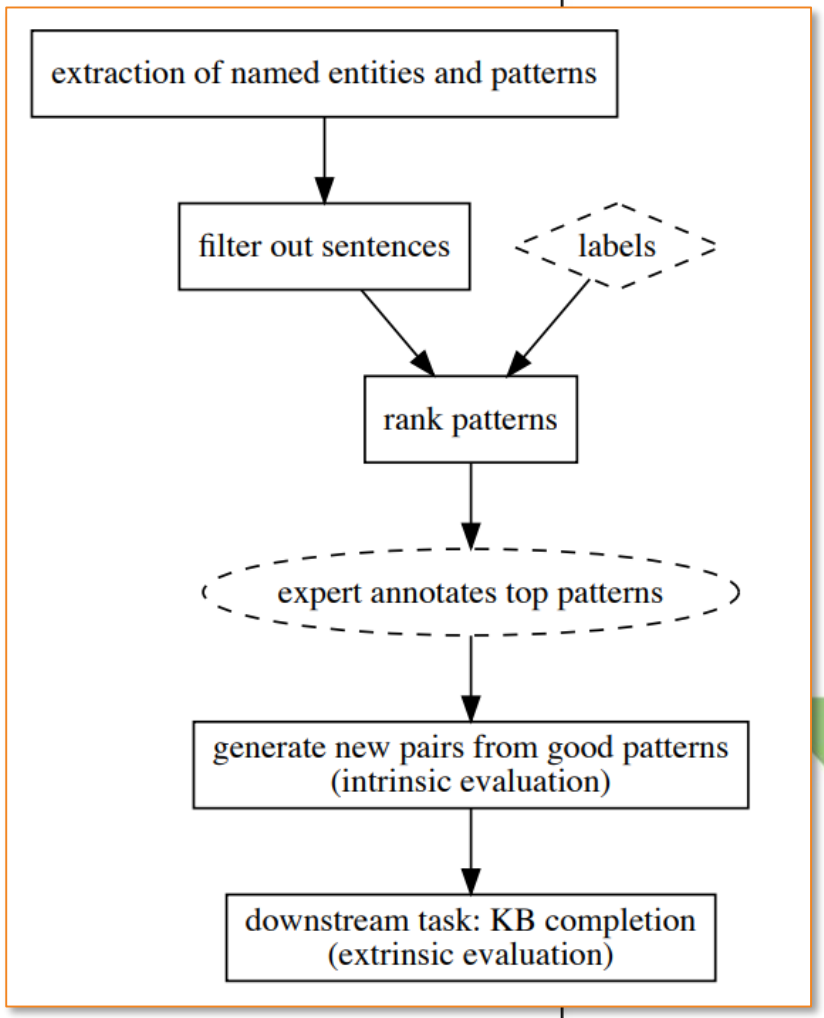


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Present: What can we do with ML in CR?



Ref. 7, Fauqueur

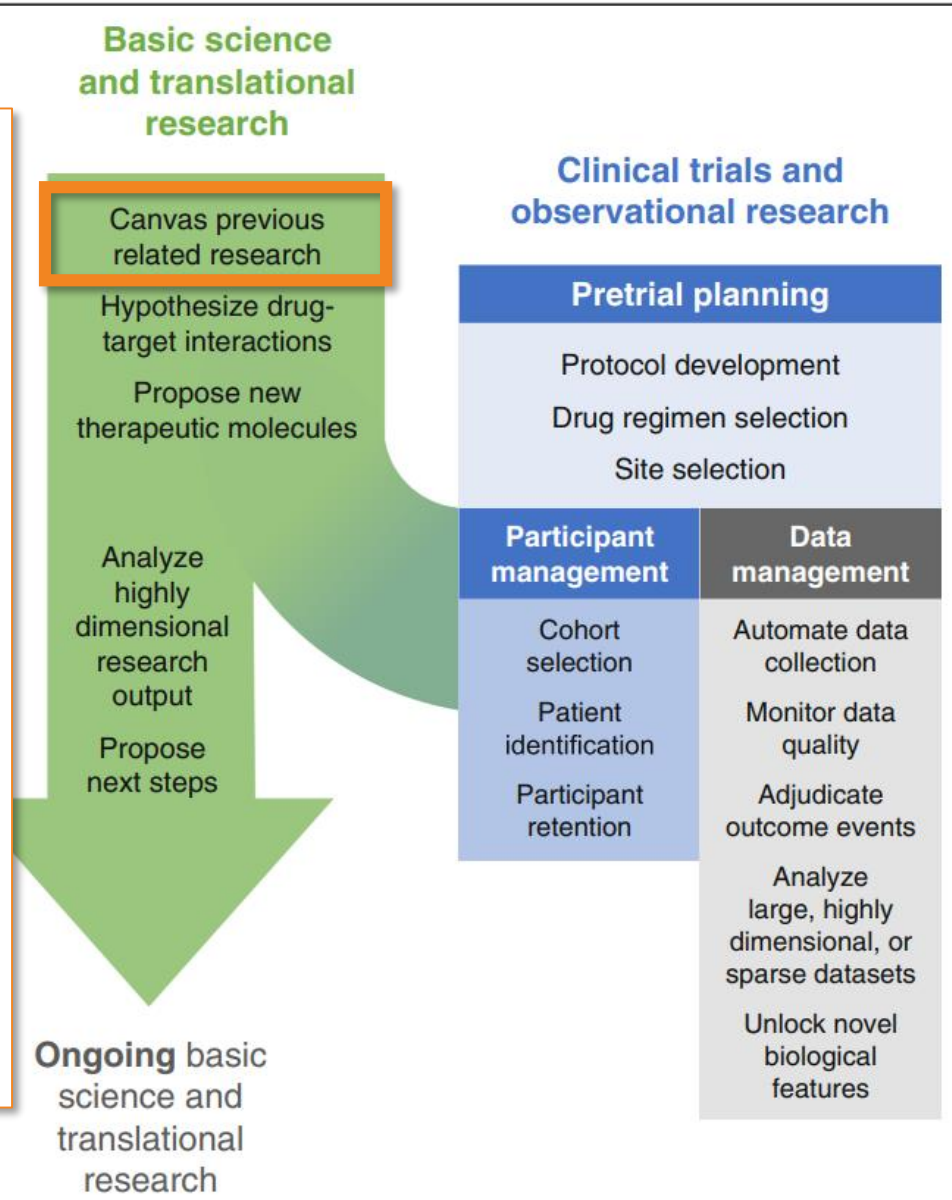


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Present: What can we do with ML in CR?

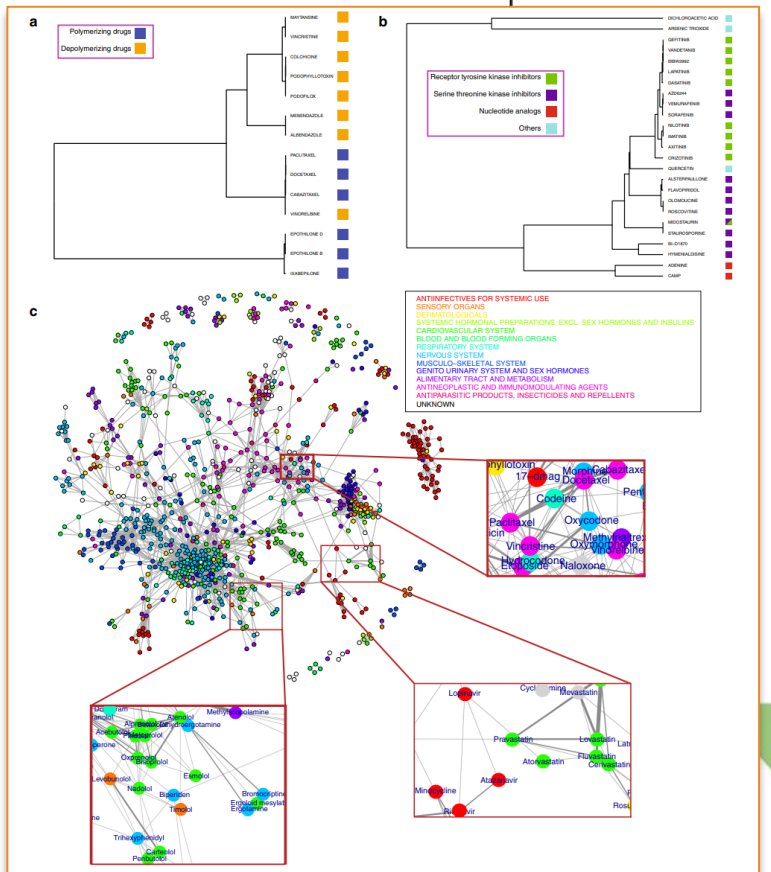


Fig. 6 BANDIT can predict specific mechanisms of action and connections between drug classes. **a** Hierarchical clustering of drugs known to target microtubules and **b** drugs known to target protein kinases. **c** Network of drugs based on shared target interactions. Drugs are colored based on their most prevalent ATC code. Three specific clusters corresponding to beta-blockers and Parkinson's medications, anti-retrovirals and statins, and opioids and antimicrotubule drugs are highlighted

Ref. 8, Madhukar

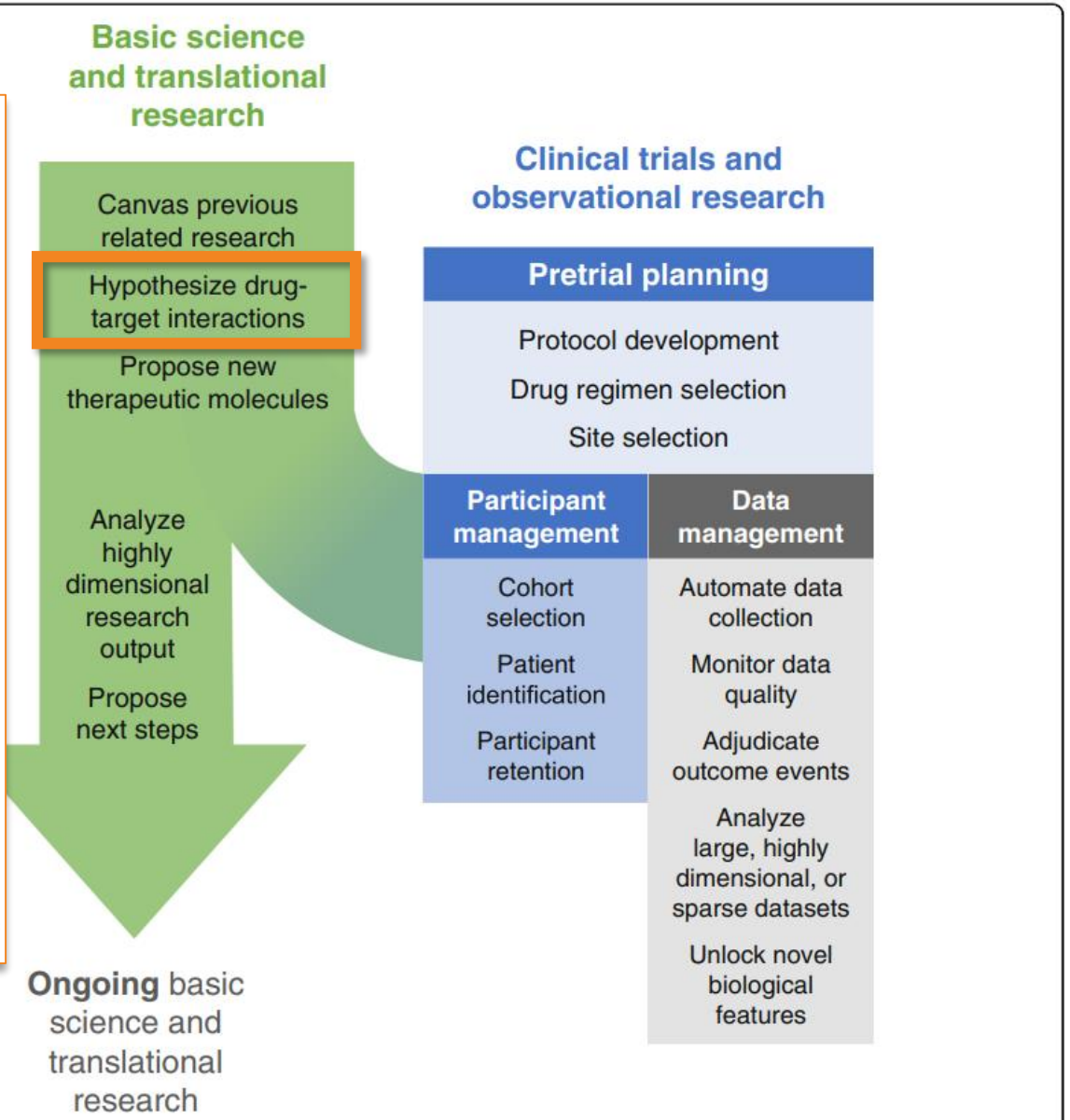


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Present: What can we do with ML in CR?

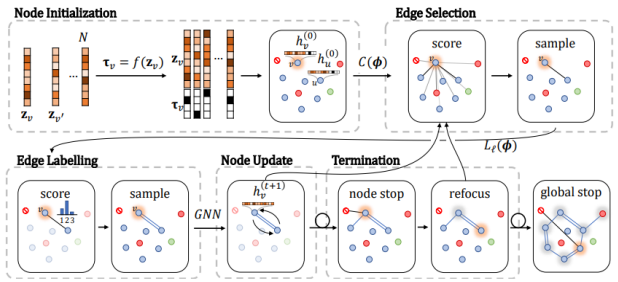


Figure 1: Illustration of the phases of the generative procedure. Nodes are initialized with latent variables and then we enter a loop between edge selection, edge labelling and node update steps until the special stop node ⊘ is selected. We then refocus to a new node or terminate if there are no candidate focus nodes in the connected component. A looped arrow indicates that several loop iterations may happen between the illustrated steps.

Ref. 9, Liu

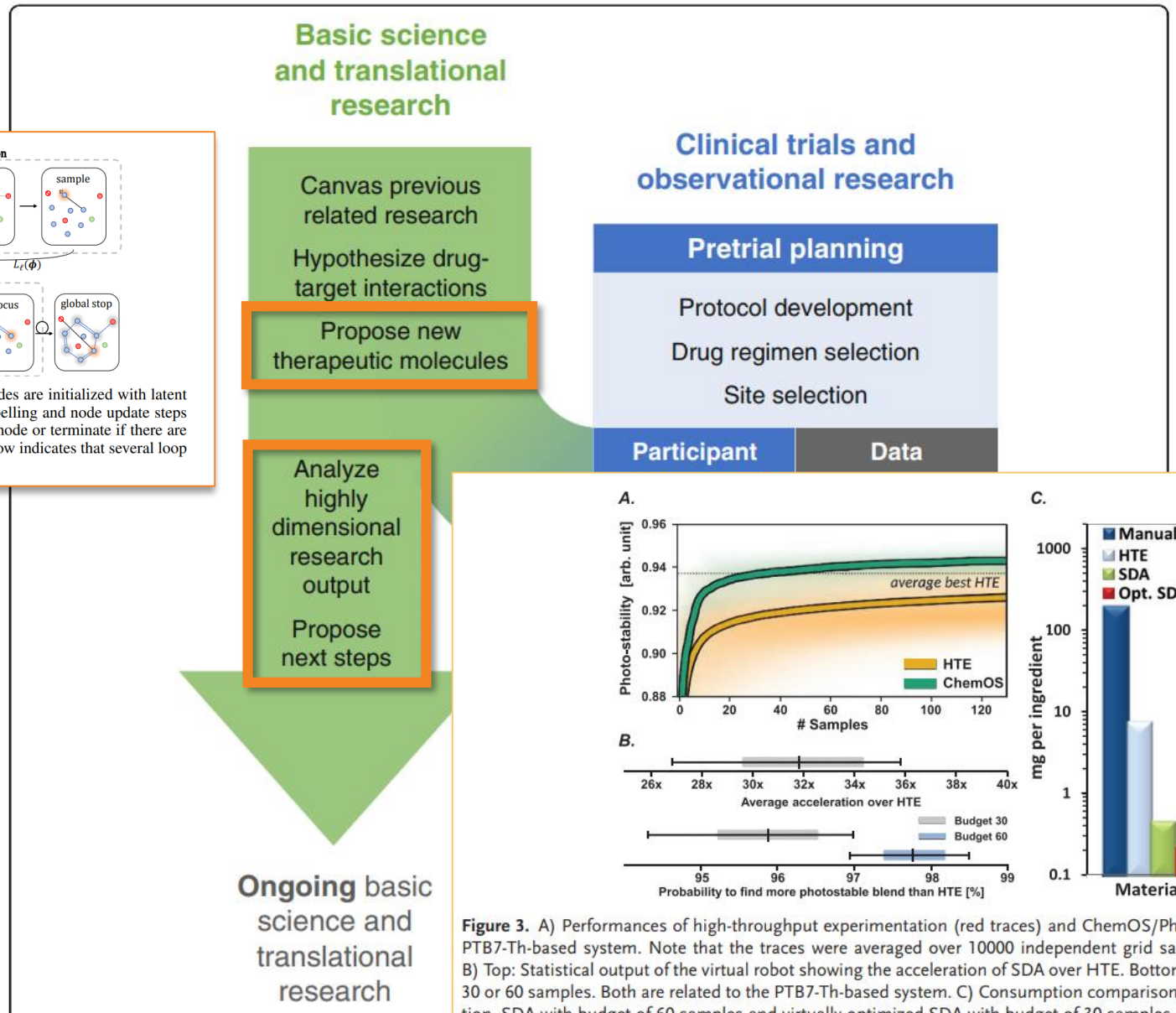


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

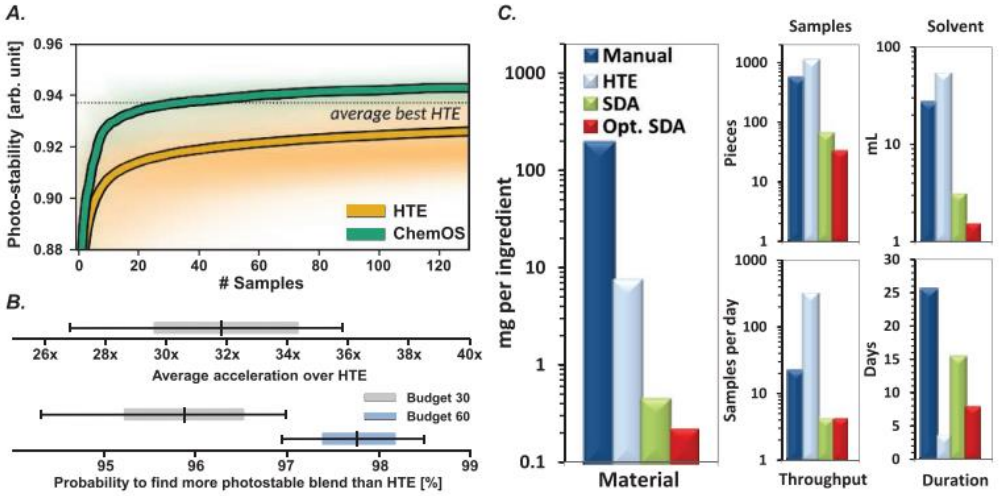


Figure 3. A) Performances of high-throughput experimentation (red traces) and ChemOS/Phoenics (blue traces) on the virtual robots related to the PTB7-Th-based system. Note that the traces were averaged over 10000 independent grid samplings and 200 independent ChemOS optimizations. B) Top: Statistical output of the virtual robot showing the acceleration of SDA over HTE. Bottom: Confidence to improve on HTE within given budget of 30 or 60 samples. Both are related to the PTB7-Th-based system. C) Consumption comparison between manual testing, high-throughput experimentation, SDA with budget of 60 samples and virtually optimized SDA with budget of 30 samples. The calculations are limited to one quaternary system.

Ref. 10, Langner

Present: What can we do with ML in CR?

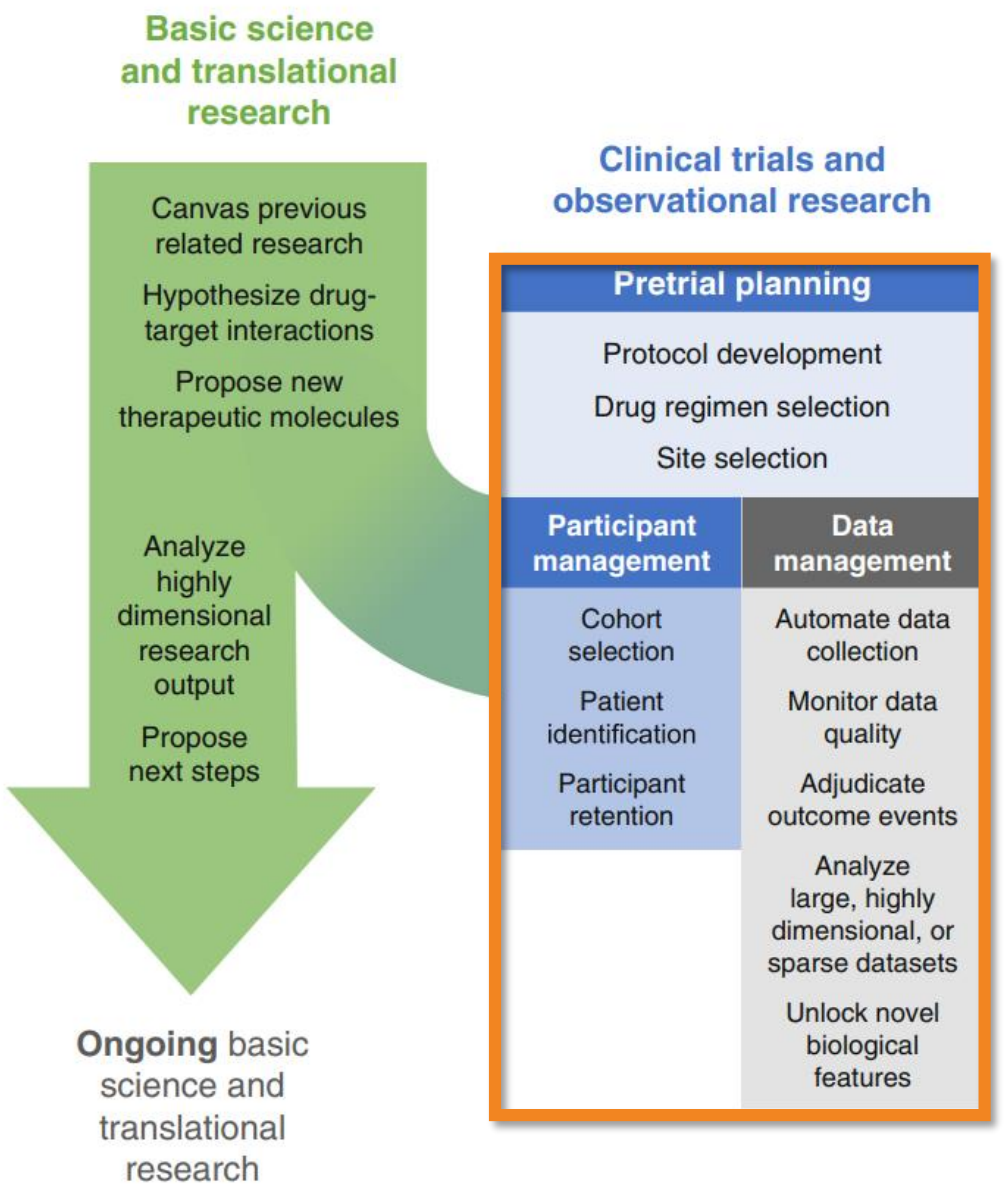
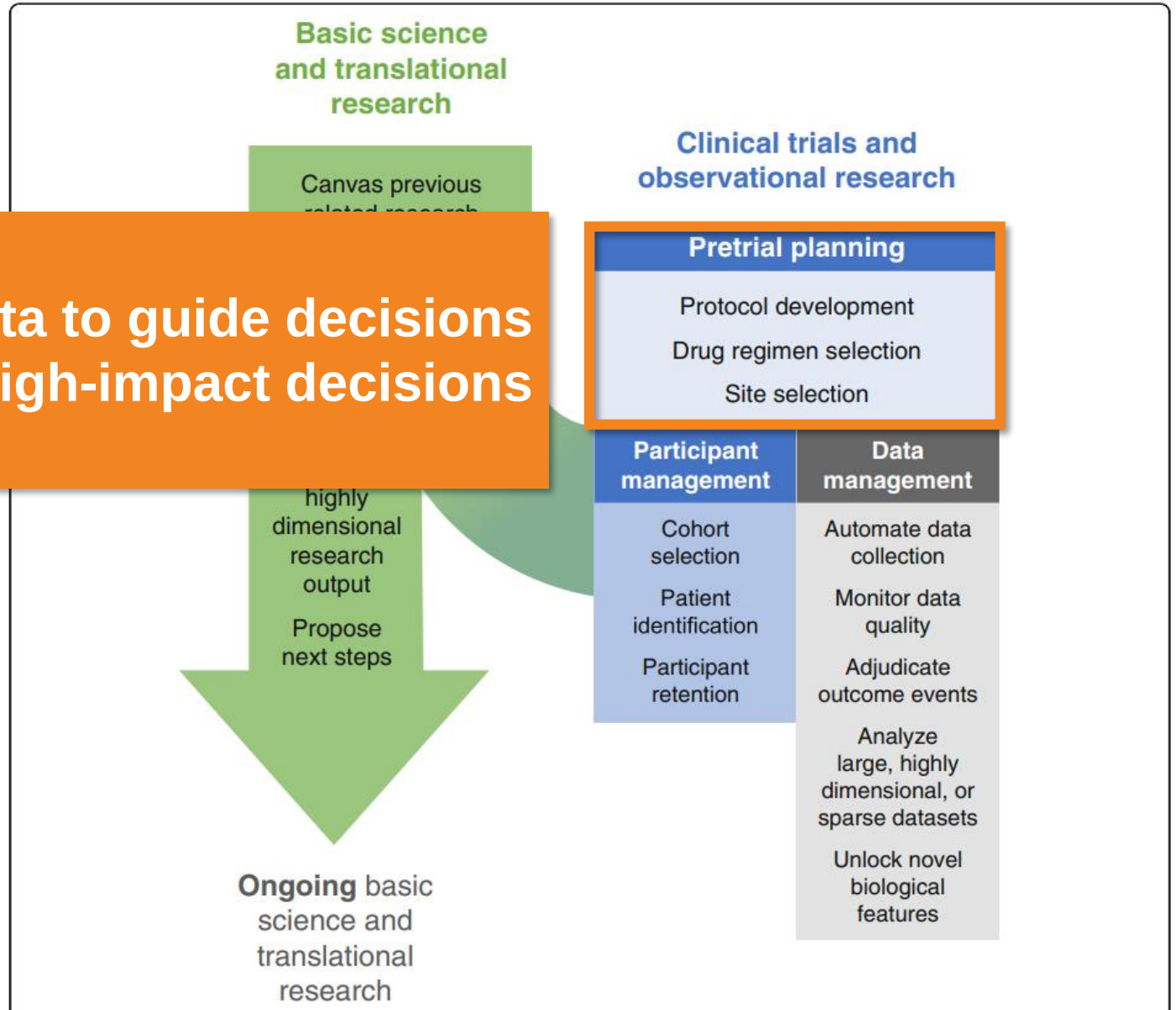


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?



Few data to guide decisions
High-impact decisions

Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

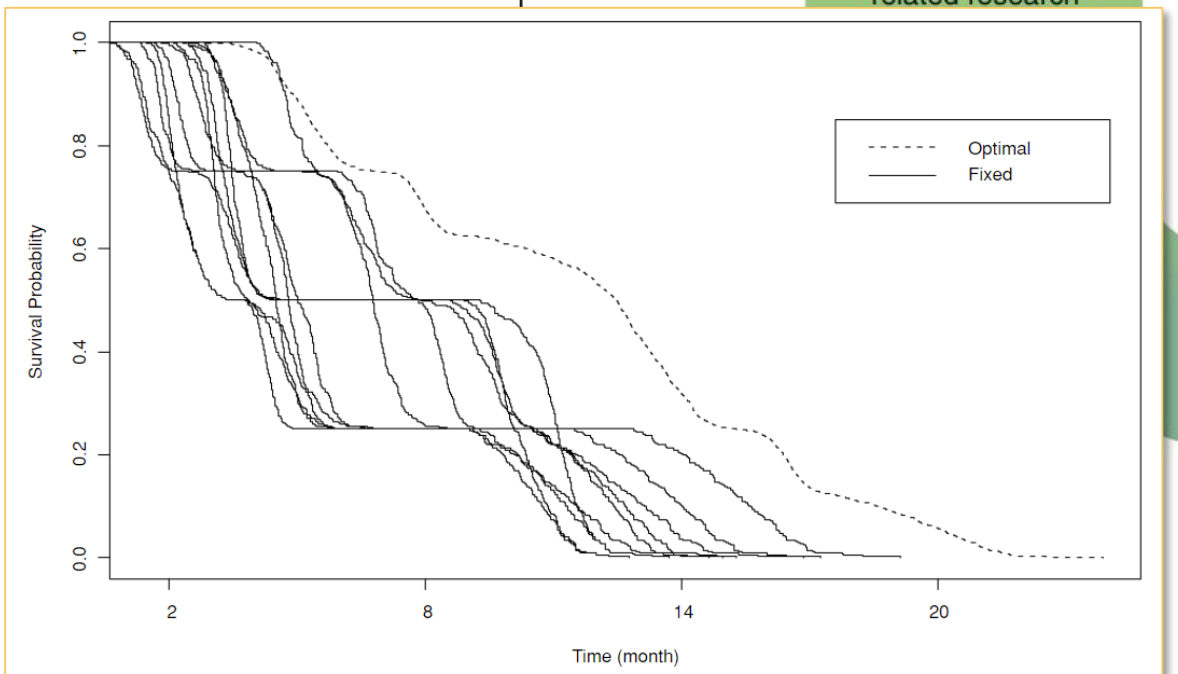


Figure 4. Survival functions for testing sample treated by 13 different regimens (12 fixed treatments plus optimal regimen).

Ref. 11, Zhao

Basic science and translational research

Canvas previous related research

Clinical trials and observational research

Pretrial planning

Protocol development

Drug regimen selection

Site selection

Participant management	Data management
Cohort selection	Automate data collection
Patient identification	Monitor data quality
Participant retention	Adjudicate outcome events
	Analyze large, highly dimensional, or sparse datasets
	Unlock novel biological features

Ongoing basic science and translational research

Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

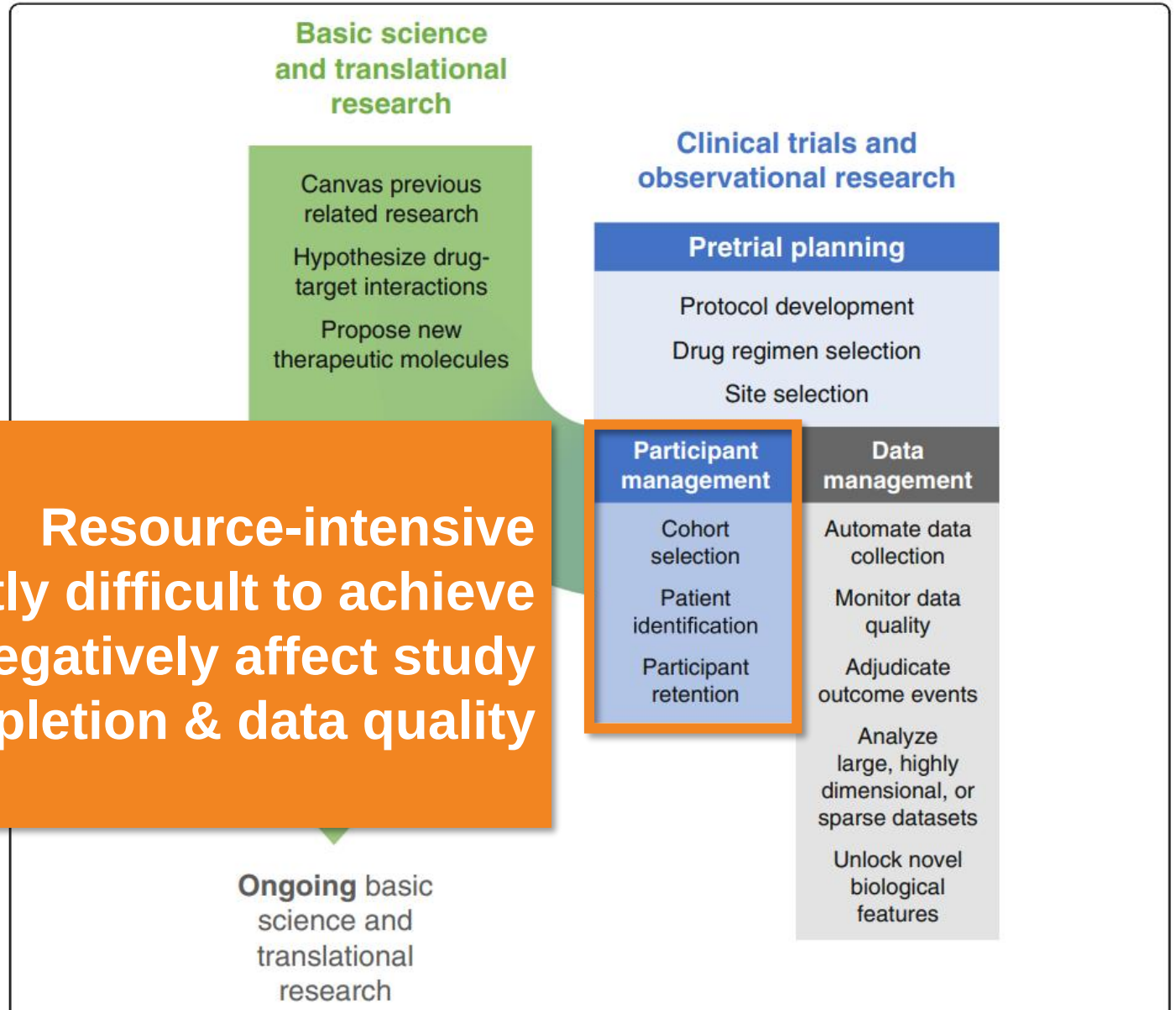


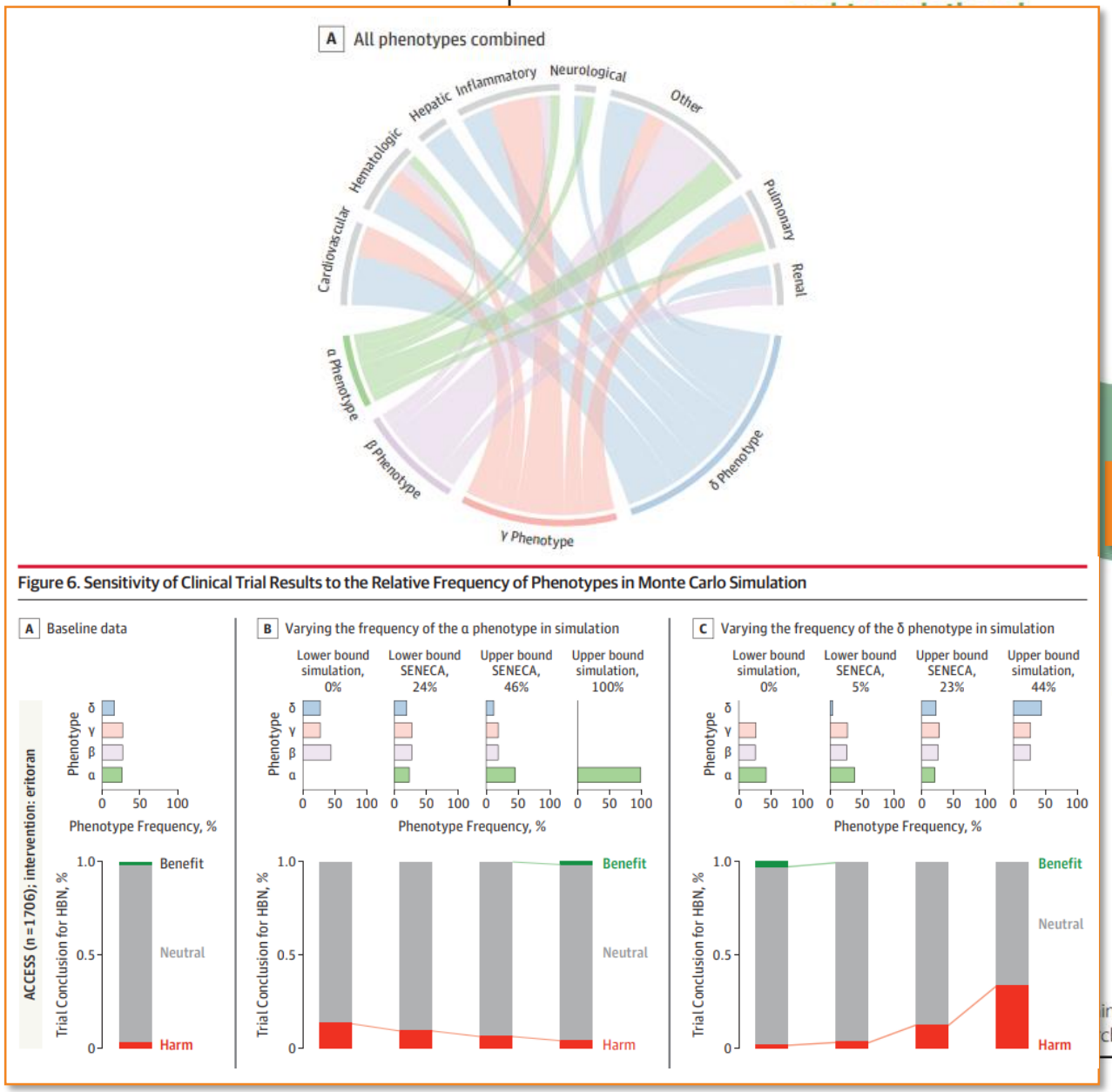
Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Resource-intensive
Persistently difficult to achieve
Can negatively affect study completion & data quality

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

Ref. 12, Seymour



Basic science

Clinical trials and observational research

Pretrial planning

- Protocol development
- Drug regimen selection
- Site selection

Participant management	Data management
Cohort selection	Automate data collection
Patient identification	Monitor data quality
Participant retention	Adjudicate outcome events
	Analyze large, highly dimensional, or sparse datasets
	Unlock novel biological features

ine learning has the potential to contribute to clinical research through
ch and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

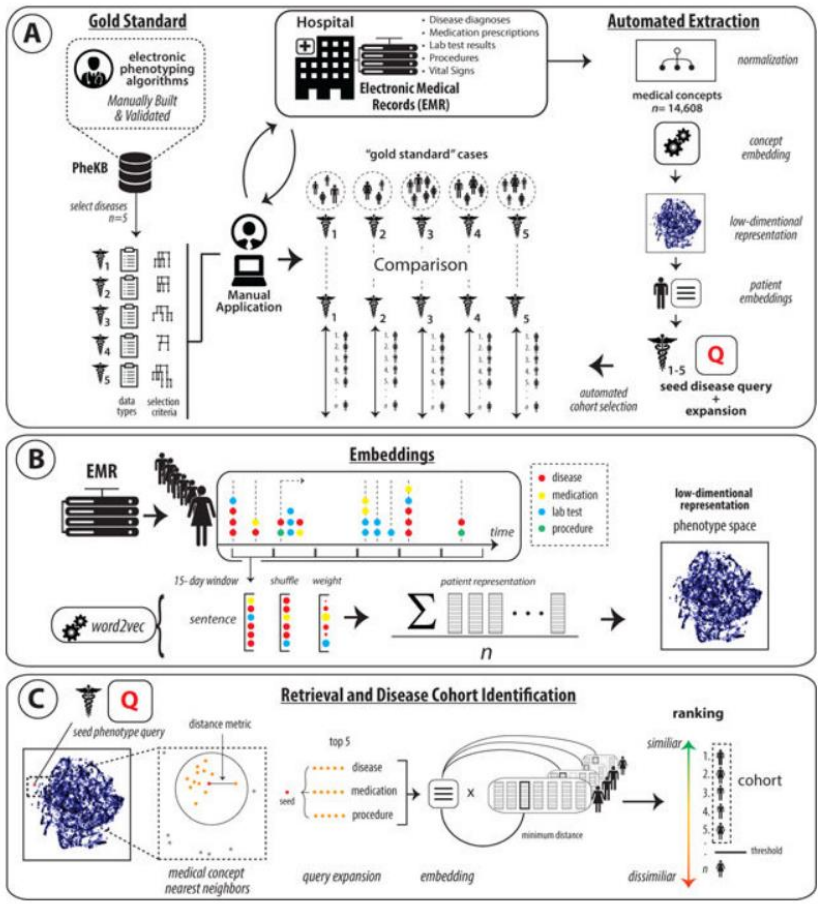


Figure 1.
Workflow design of the current study. A) Framework for comparing gold standard PheKB algorithms to our automated method. B) Embeddings procedure. C) Retrieval and disease cohort identification

Ref. 13, Glicksberg

Basic science and translational research

Clinical trials and observational research

Pretrial planning

Protocol development

Drug regimen selection

Site selection

Participant management	Data management
Cohort selection	Automate data collection
Patient identification	Monitor data quality
Participant retention	Adjudicate outcome events
	Analyze large, highly dimensional, or sparse datasets
	Unlock novel biological features

Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

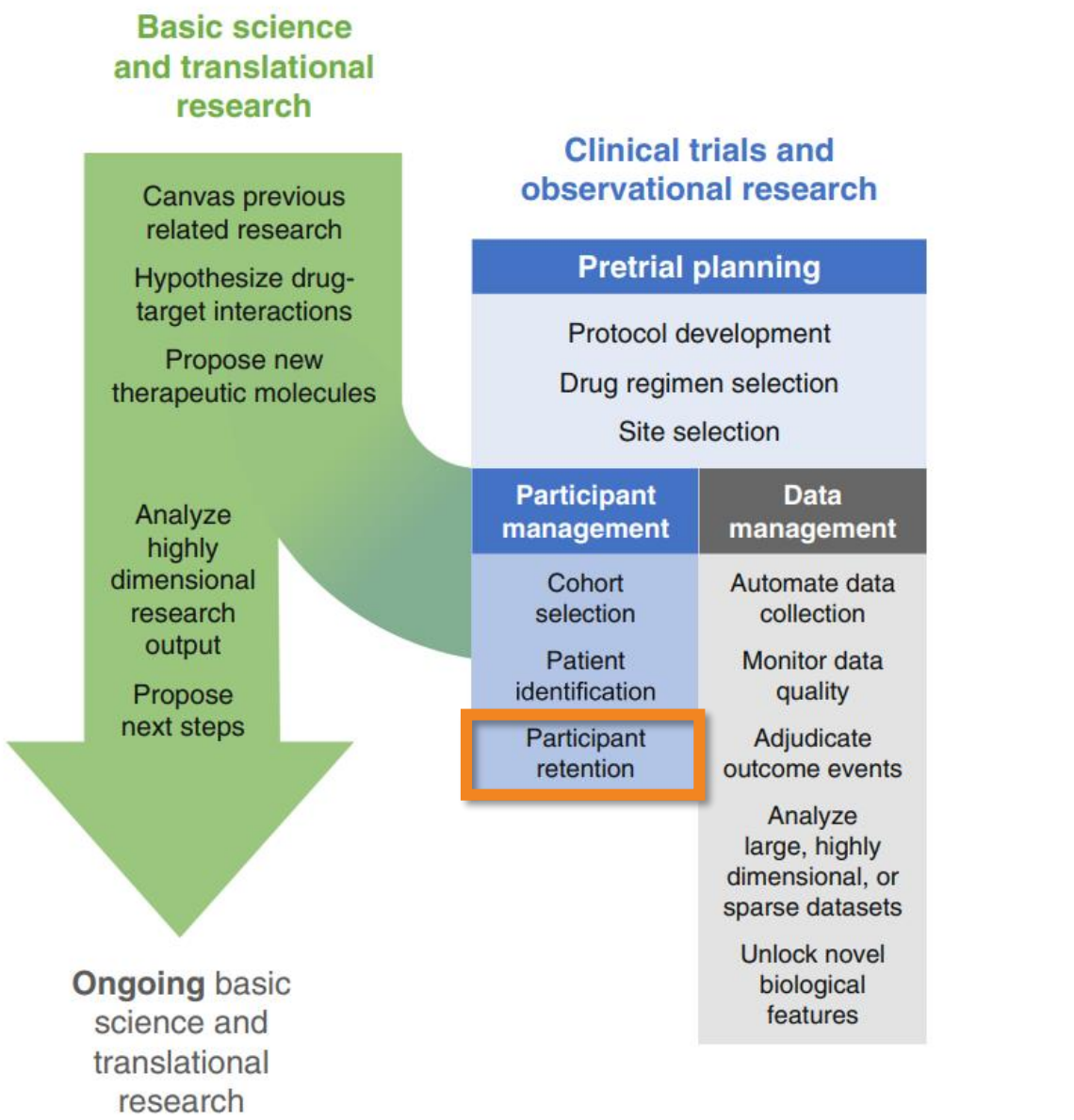


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?

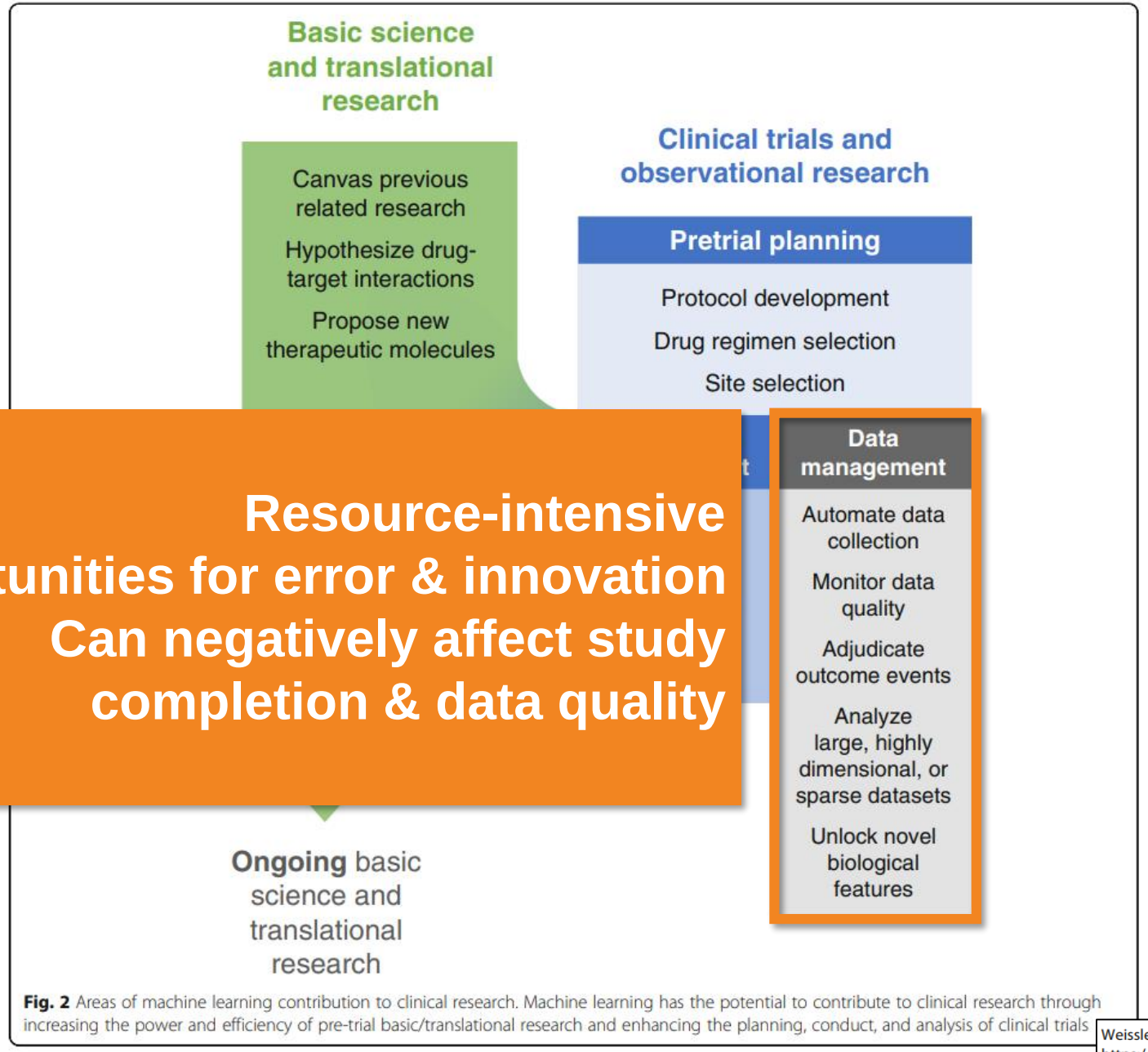
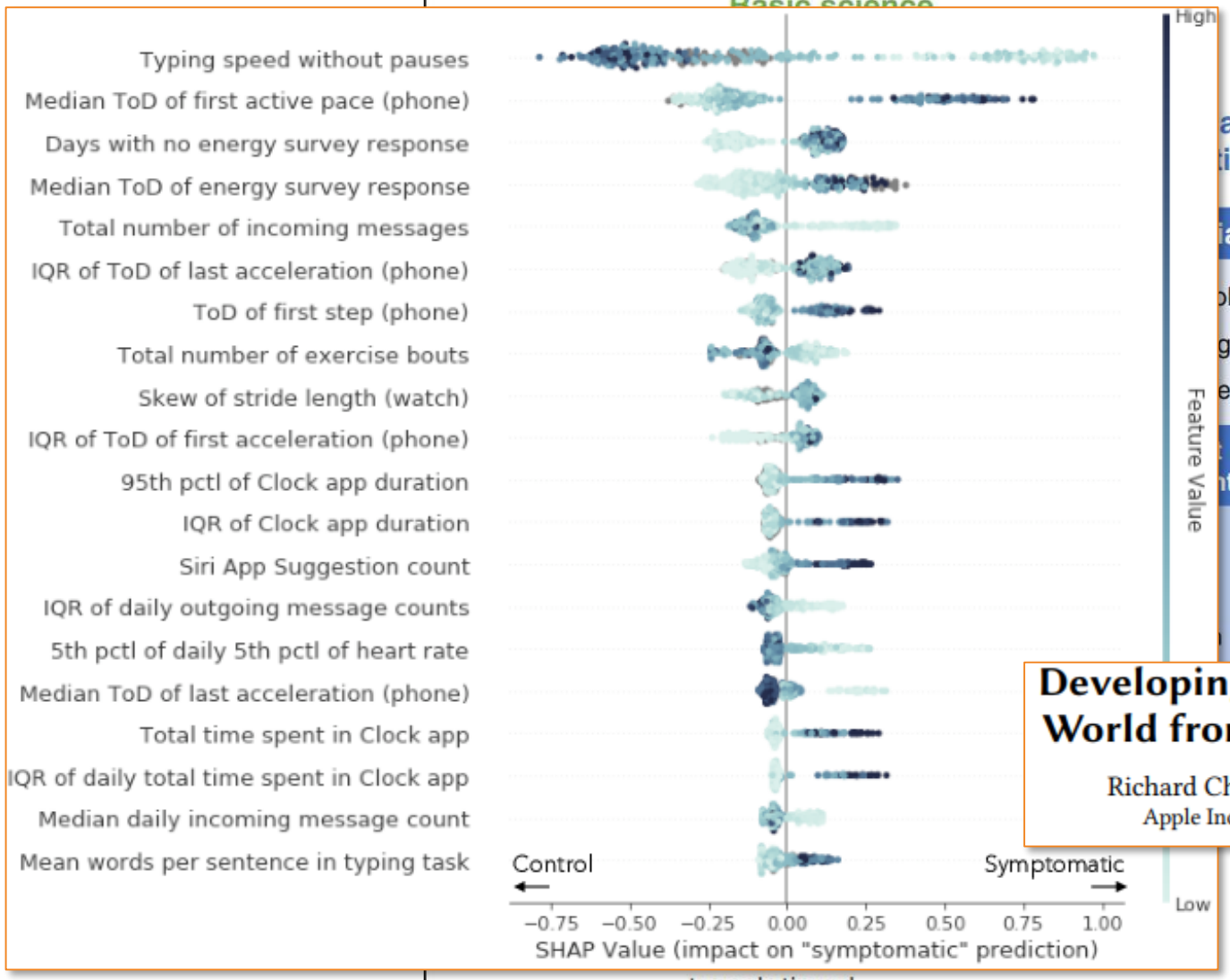


Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Resource-intensive
Opportunities for error & innovation
Can negatively affect study completion & data quality

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?



Ref. 14, Chen

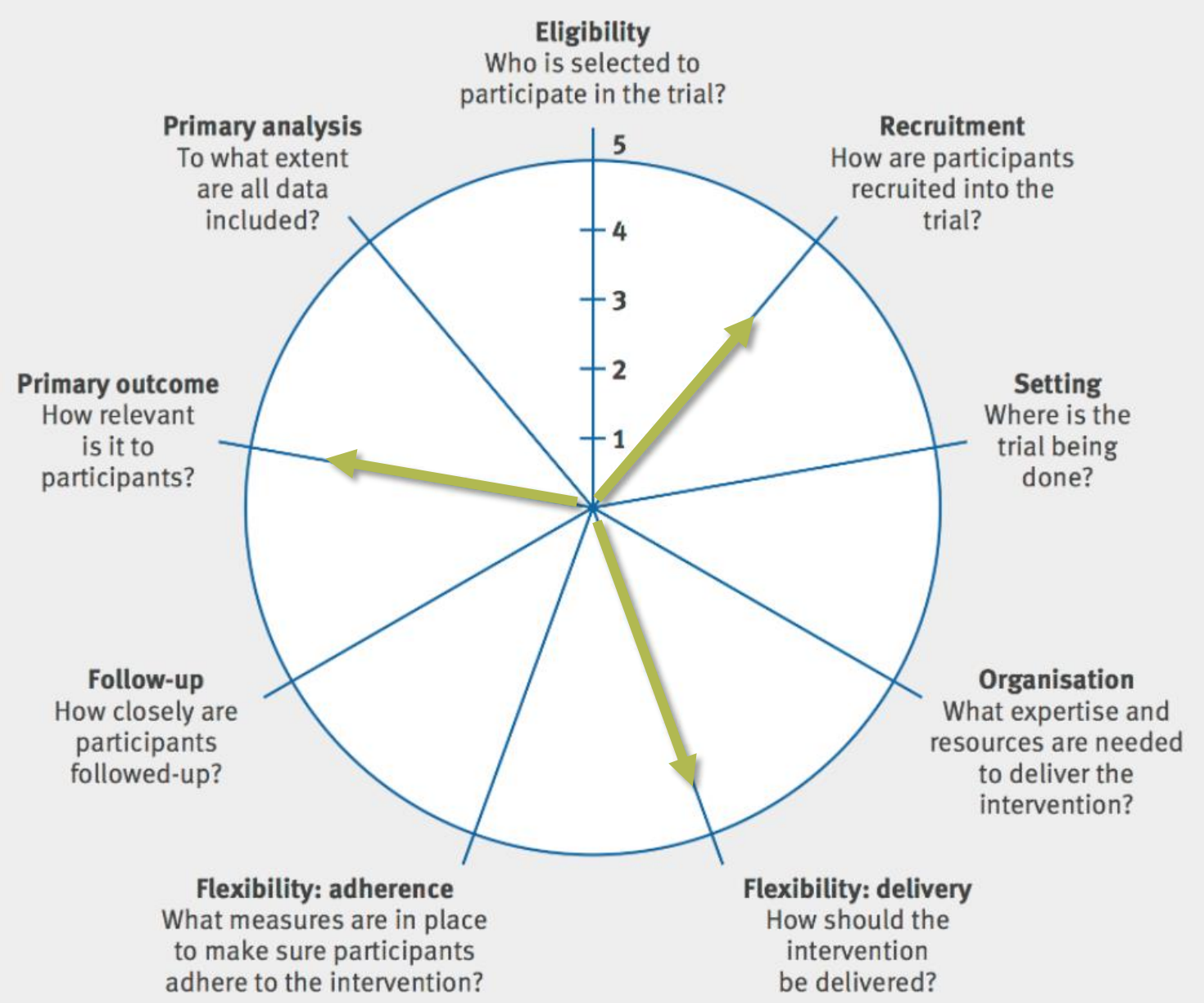
Fig. 2 Areas of machine learning contribution to clinical research. Machine learning has the potential to contribute to clinical research through increasing the power and efficiency of pre-trial basic/translational research and enhancing the planning, conduct, and analysis of clinical trials

Developing Measures of Cognitive Impairment in the Real World from Consumer-Grade Multimodal Sensor Streams

Richard Chen* Apple Inc. | Filip Jankovic* Evidation Health, Inc. | Nikki Marinsek* Evidation Health, Inc.

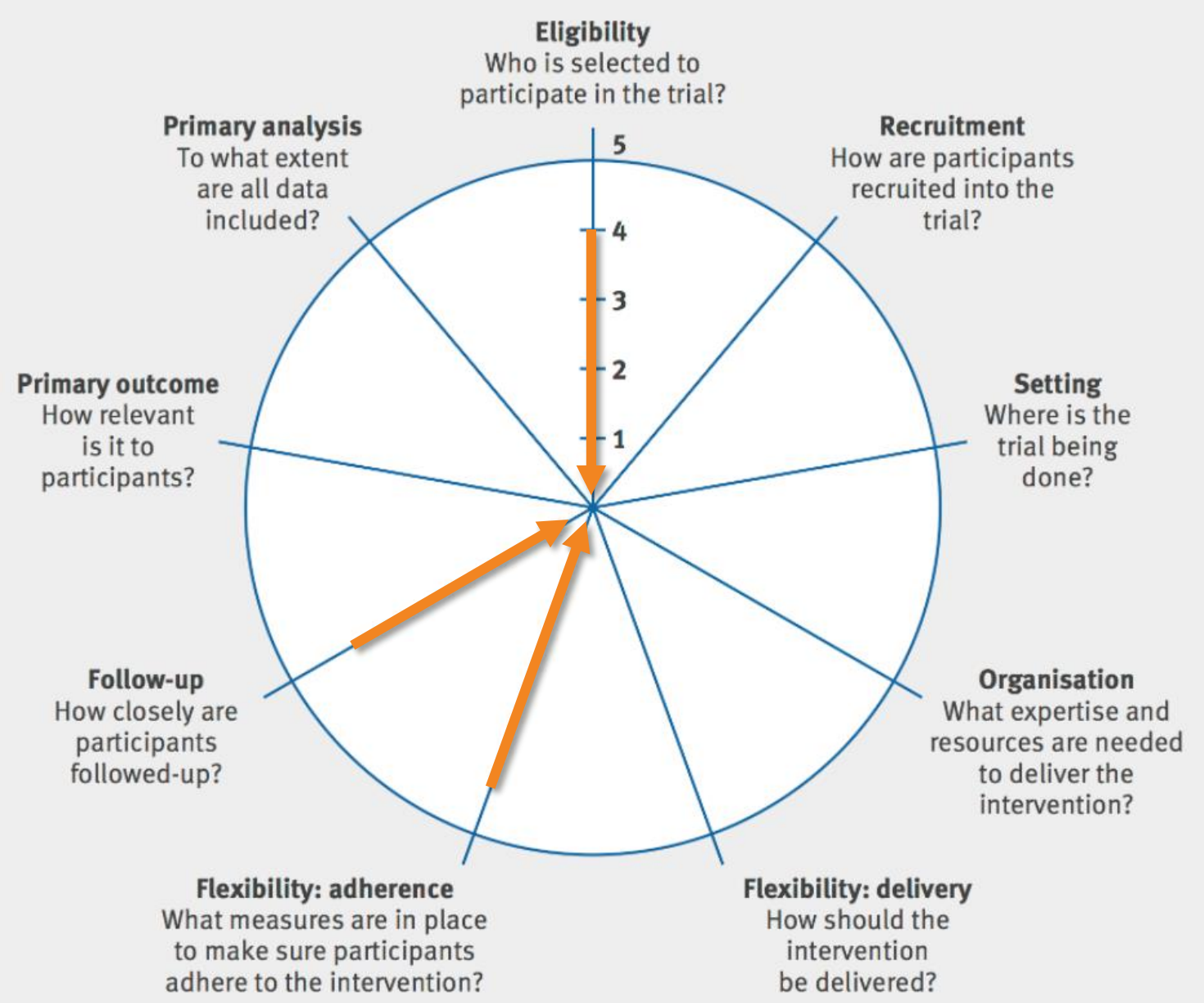
Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Present: What can we do with ML in CR?



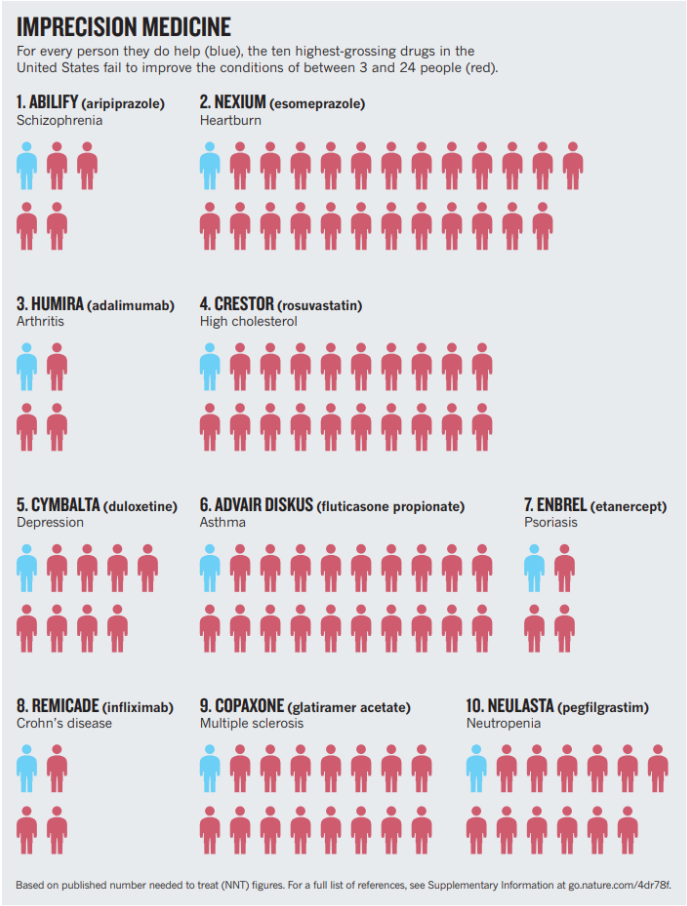
Ref. 14, Loudon

Present: What can we do with ML in CR?



Ref. 14, Loudon

Present: What's stopping us from using ML in CR?



Ref. 16, Schork

Ethical risk

Ethical imperative

Present: What's stopping us from using ML in CR?

Defining the Potential of Machine Learning and Artificial Intelligence Approaches	Erich Huang, *Duke Clinical Research Institut* Marzyeh Ghassemi, *University of Toronto*
AI/ML Approaches to Enable Healthcare Delivery vs. Clinical Research	Erich Huang, *Duke Clinical Research Institut*
AI/ML Framework for Regulated Clinical Research - Opening Thoughts	Marzyeh Ghassemi, *University of Toronto*
Biomarker Discoveries and Drug Target Optimization	James Benoit, *Harvard University* Daniel Freitag, *Bayer* Paul Slater, *Microsoft*
Cohort Composition/Phenotyping and Patient Identification	Scott Kollins, *Duke University* Shameer Khader, *AstraZeneca* Tristan Naumann, *Microsoft*

Transforming Data Surveillance During Trial Conduct	Matthew Roe, *Duke Clinical Research Institute* Zhaoling Meng, *Sanofi* Ricardo Henao, *Duke University*
Optimization of Trial Operational Conduct	Erich Huang, *Duke Clinical Research Institute* Mohanish Anand, *Pfizer* Lucas Glass, *IQVIA* Bram Zuckerman, *FDA*
– Monitoring of Patient Compliance and Adherence	Yuan Luo, *Northwestern University*
– Advanced Risk-Based Monitoring Techniques	Erich Huang, *Duke Clinical Research Institute*
– Endpoint Adjudication	Emmette Hutchison, *AstraZeneca*
State of the Art Methods Talk	Marzyeh Ghassemi, *University of Toronto*

What **can** we do?

What **should** we do and how?

Common Access to Data Sets	Yuan Luo, *Northwestern University* Marzyeh Ghassemi, *University of Toronto* Jeff Riesmeyer, *Eli Lilly* Matthew Diamond, *FDA* Paul Varghese, *Verily*
3rd Party Verification and Certification for Algorithms	Yuan Luo, *Northwestern University* Stephen Browning, *FDA* Ricardo Henao, *Duke University*
Communities, Challenges, Common Focus Areas	Brian Bot, *Sage Bionetworks* Masahiro Murakami, *Eli Lilly* Bray Patrick-Lake, *Evidation*

Broader Regulatory Perspective on the Applications of AI/ML for Clinical Trials	Khair ElZarrad, *FDA*
Aligning Expectations for Methods	Olivier Elemento, *Weill Cornell Medicine*
Data Needs and Requirements to Optimize and Validate AI/ML Algorithms	Robert Ball, *FDA* Adarsh Subbaswamy, *John Hopkins University* Eamon Caddigan, *Evidation*
Expectations for Validation of Algorithms	Stephen Browning, *FDA* Brian Bot, *Sage Bionetworks* Rajesh Ranganath, *New York University*
Preferred Quality Metrics	Boris Brodsky, *FDA* Michael Hughes, *Tufts University* Philip Sarocco, *Cytokinetics* Paul Slater, *Microsoft*

Present: What’s stopping us from using ML in CR?

Operational barriers

1. Adequately skilled teams
2. Data:
   - Adequate quantity
   - Multiple sources
   - Adequate quality

Present: What's stopping us from using ML in CR?

Operational barriers

1. Adequately skilled teams
2. Data:
   - Adequate quantity
   - Multiple sources
   - Adequate quality

Philosophical barriers

1. Explainability versus trustworthiness
2. Error and bias

Present: What's stopping us from using ML in CR?

Operational barriers

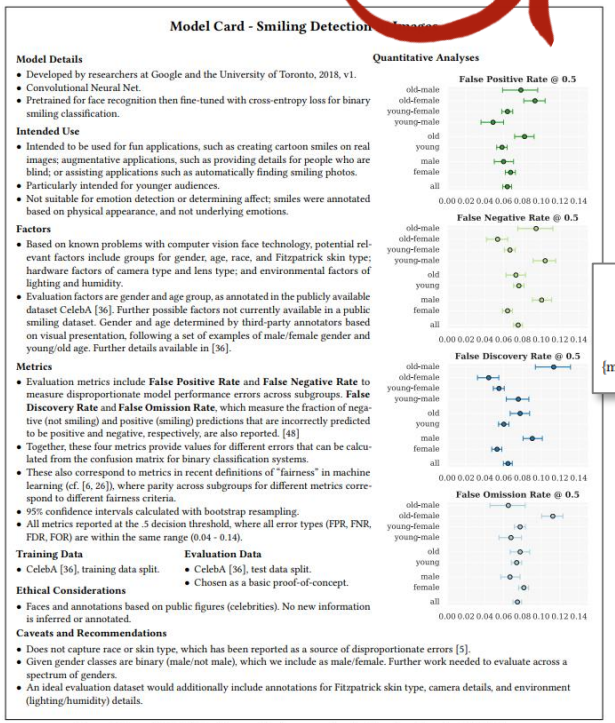
1. Adequately skilled teams
2. Data:
   - Adequate quantity
   - Multiple sources
   - Adequate quality

Validation

Reporting

Philosophical barriers

1. Explainability versus trustworthiness
2. Error and bias



Model Card - Smiling Detection in Images

Model Details
- Developed by researchers at Google and the University of Toronto, 2018, v1.
- Convolutional Neural Net.
- Pretrained for face recognition then fine-tuned with cross-entropy loss for binary smiling classification.

Intended Use
- Intended to be used for fun applications, such as creating cartoon smiles on real images; augmentative applications, such as providing details for people who are blind; or assisting applications such as automatically finding smiling photos.
- Particularly intended for younger audiences.
- Not suitable for emotion detection or determining affect; smiles were annotated based on physical appearance, and not underlying emotions.

Factors
- Based on known problems with computer vision face technology, potential relevant factors include groups for gender, age, race, and Fitzpatrick skin type; hardware factors of camera type and lens type; and environmental factors of lighting and humidity.
- Evaluation factors are gender and age group, as annotated in the publicly available dataset CelebA [36]. Further possible factors not currently available in a public smiling dataset. Gender and age determined by third-party annotators based on visual presentation, following a set of examples of male/female gender and young/old age. Further details available in [36].

Metrics
- Evaluation metrics include **False Positive Rate** and **False Negative Rate** to measure disproportionate model performance errors across subgroups. **False Discovery Rate** and **False Omission Rate**, which measure the fraction of negative (not smiling) and positive (smiling) predictions that are incorrectly predicted to be positive and negative, respectively, are also reported. [48]
- Together, these four metrics provide values for different errors that can be calculated from the confusion matrix for binary classification systems.
- These also correspond to metrics in recent definitions of "fairness" in machine learning (cf. [6, 26]), where parity across subgroups for different metrics correspond to different fairness criteria.
- 95% confidence intervals calculated with bootstrap resampling.
- All metrics reported at the .5 decision threshold, where all error types (FPR, FNR, FDR, FOR) are within the same range (0.04 - 0.14).

Training Data
- CelebA [36], training data split.

Evaluation Data
- CelebA [36], test data split.
- Chosen as a basic proof-of-concept.

Ethical Considerations
- Faces and annotations based on public figures (celebrities). No new information is inferred or annotated.

Caveats and Recommendations
- Does not capture race or skin type, which has been reported as a source of disproportionate errors [5].
- Given gender classes are binary (male/not male), which we include as male/female. Further work needed to evaluate across a spectrum of genders.
- An ideal evaluation dataset would additionally include annotations for Fitzpatrick skin type, camera details, and environment (lighting/humidity) details.

Quantitative Analyses

Figure 2: Example Model Card for a smile detector trained and evaluated on the CelebA dataset.

Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru
{mmitchellai,simonewu,andrewzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com
deborah.raji@mail.utoronto.ca

Ref. 17

Present: What's stopping us from using ML in CR?

trials.ai. Accessed February 2, 2021.

trials.ai

ABOUT US LAB NOTES CONTACT US

Our Smart Protocol system aims to automate and add intelligence to clinical trial design with AI

Clinical trials and observational research

Pretrial planning

Protocol development

Drug regimen selection

Site selection

Participant management

Cohort selection

Patient identification

Participant retention

Data managem

Monitor da

https://deep6.ai/how-it-works/

DEEP6 AI

HOW IT WORKS SCHEDULE A DEMO ABOUT US CAREERS BLOG CONTACT

FIND MORE PATIENTS IN MINUTES NOT MONTHS

BULLFROGAI

HOME TECHNOLOGY PARTNERING SOLUTIONS PIPELINE LEADERSHIP INVESTORS MEDIA

DISCOVERING THERAPEUTICS THAT WOULDN'T BE FOUND BEFORE.

88	8½	2.6	100	204
PERCENT OF DRUGS IN PIPELINE WILL FAIL	YEARS AVERAGE UNTIL READY FOR MARKET	BILLION AVERAGE COST FOR NEW DRUG	BILLION SPENT ON 208 RESEARCH BY TOP 15 PHARMA	BILLION TOTAL ANNUAL R&D SPENDING BY 2024; 3% CAGR

. https://www.bullfrogai.com/our-solution/

aicure.com

AiCure

EVENTS & NEWS BLOG LEARN OPENDBM PARTNERSHIPS COMPANY CONTACT

THE RIGHT DOSE FOR THE RIGHT PATIENT

AiCure delivers a compliant, scalable AI platform to maximize the impact of data on clinical research and end-to-end operations, from pre-clinical to commercialization.

AiCure brings AI to the life sciences industry to guide data-driven decision-making for more meaningful clinical trials, optimized drug development, and improved business operations.

Fig. 2

increas

tial to contribute to clinical research through

nning, conduct, and analysis of clinical trials

MENDEL

Products

Recruit

Increase Enrollment by 50%. Prescreen in Under 10 Minutes.

Recruit is an AI-powered solution for clinical sites looking to accelerate patient prescreening,

Mendel.ai

Present: What's stopping us from using ML in CR?

Med Health Care and Philos (2016) 19:177–190
DOI 10.1007/s11019-015-9661-6

CrossMark

SCIENTIFIC CONTRIBUTION

"You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research

Sigrid Sterckx[1] · Vojin Rakic[2] · Julian Cockbain[3] · Pascal Borry[4]

Ref. 18

SPECIAL REPORT

An invisible hand: Patients aren't being told about the AI systems advising their care

By Rebecca Robbins and Erin Brodwin July 15, 2020

Reprints

Ref. 19

Future: Overcoming barriers to implementation

→ Erich & a discussion of data

Data Liquidity

what is it really?

Why Data Liquidity ?

Why?

LEARNING HEALTH requires unceasing data collection & knowledge generation that is plowed back into patient care

INTERPRET

ANALYZE

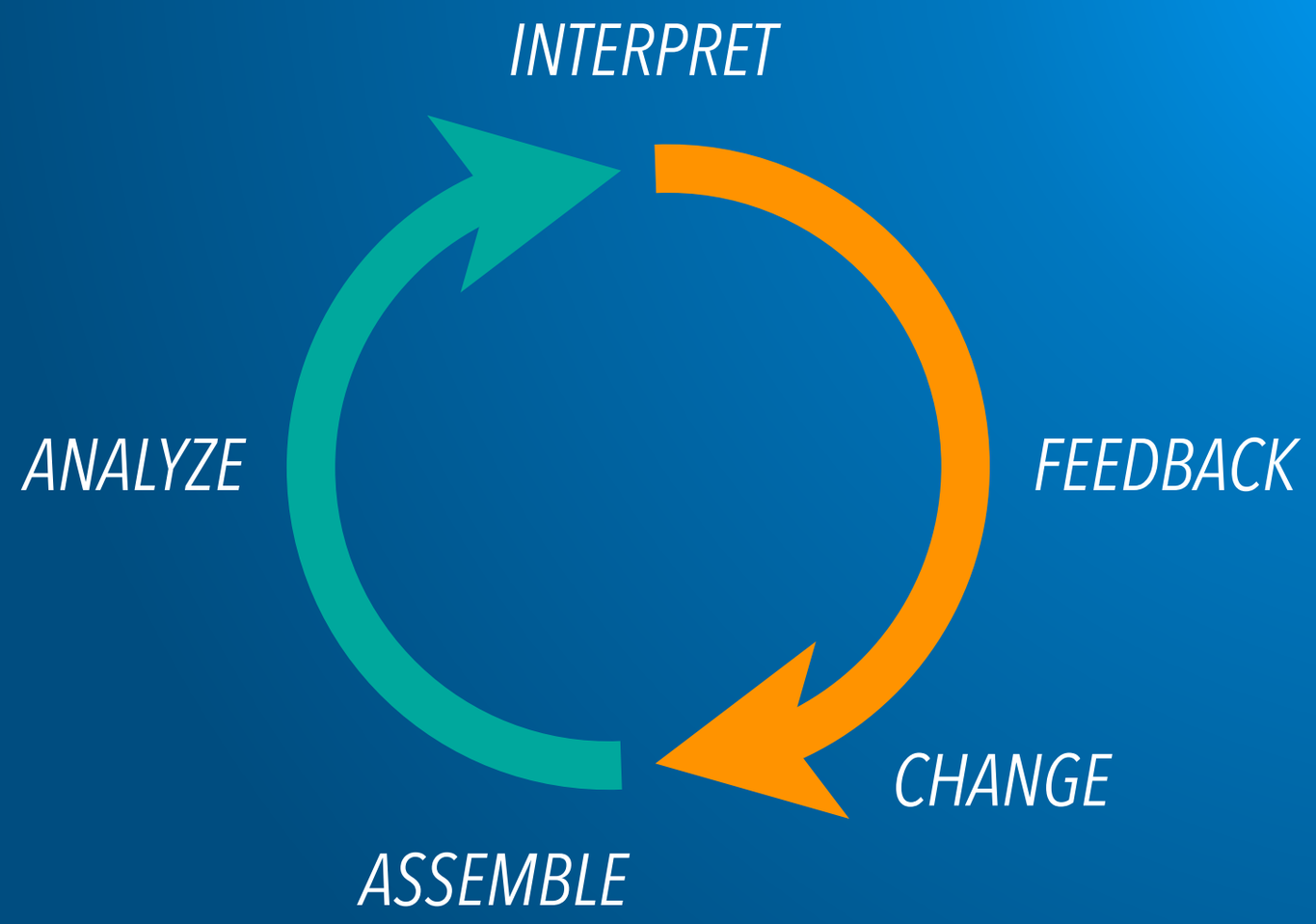
FEEDBACK

CHANGE

ASSEMBLE

Why?

LEARNING HEALTH requires unceasing data collection & knowledge generation that is plowed back into patient care



A Useful Analogy

Analogy

Analogy

If you can go to an ATM in Antwerp, or anywhere, you can securely access your $ in your US bank with virtually no friction other than a fee

Analogy

Analogy

Analogy

Analogy

Common stock issued by
B&O Railroad Co. in 1903

Mortgage bond issued by Cleveland
Short-Line Railway Co. in 1911

Promissory note issued by the
Imperial Bank of India in 1926

Promissory note issued by the 2nd
Bank of the United States in 1840

Common stock issued by
Pennsylvania Railroad Co. in 1959

Common stock issued by
Reading Co. in 1969

Mortgage note issued (signed) by
"Shoeless"Joe Jackson in 1941

Certificate of deposit issued by the
U.S. Postal Savings System in 1932

Federal reserve note issued by
the U.S. Government in 2009

Data Liquidity

Data Liquidity

Data Liquidity

Data Liquidity

Data Liquidity

Data Liquidity

Data Liquidity

" I don't know how many aortic stenosis patients I have "

Currently, it's difficult to even obtain basic counts of patients…

Data Liquidity

" I don't know how many prostate cancer patients I have "

Currently, it's difficult to even obtain basic counts of patients…

Data Liquidity

Beyond counts, what does data liquidity look like?

Data Liquidity

Beyond counts, what does data liquidity look like?

How many patients meet the eligibility criteria for this RSV study?

What is the door-to-balloon time for the past month?

How many MACE events did we see for this cohort in the last 6 months?

Data Liquidity

Beyond counts, what does data liquidity look like?

How many patients meet the eligibility criteria for this RSV study?

What is the door-to-balloon time for the past month?

How many MACE events did we see for this cohort in the last 6 months?

I want to join liver MRIs with radiology and pathology reports for deep learning to predict hepatic cancer outcomes

Data Liquidity

Beyond counts, what does data liquidity look like?

How many patients meet the eligibility criteria for this RSV study?

What is the door-to-balloon time for the past month?

How many MACE events did we see for this cohort in the last 6 months?

I want to join liver MRIs with radiology and pathology reports for deep learning to predict hepatic cancer outcomes

An answer—with appropriate security and compliance controls—should be obtainable in minutes

Data Liquidity

So what might liquidity look like?

```
import urllib
import json
import time
from twilio.rest import TwilioRestClient

# Twilio info
account_sid = "ACda2817bd10290bff2a66021d6204c94a"
auth_token  = "<SECRET DATA>"
from_phone = "+13107517490"
to_phone = "+18182937524"
client = TwilioRestClient(account_sid, auth_token)

# How far back we want to look for results
start_time = str(time.time() - 60 * 60) # An hour ago

# Practioner to fetch labs for
practioner_id = 259386

# URL to fetch lab results from
url="".join(["https://api.liquidity.duhs.com?_format=json&preformer=",
practioner_id, "&issued=>", start_time])
results = json.loads(urllib.urlopen(url));

# Loop through all results in the past hour to see if they are normal or not
for result in results:
  if data["interpretation"]["coding"][0]["code"] != "N":
```

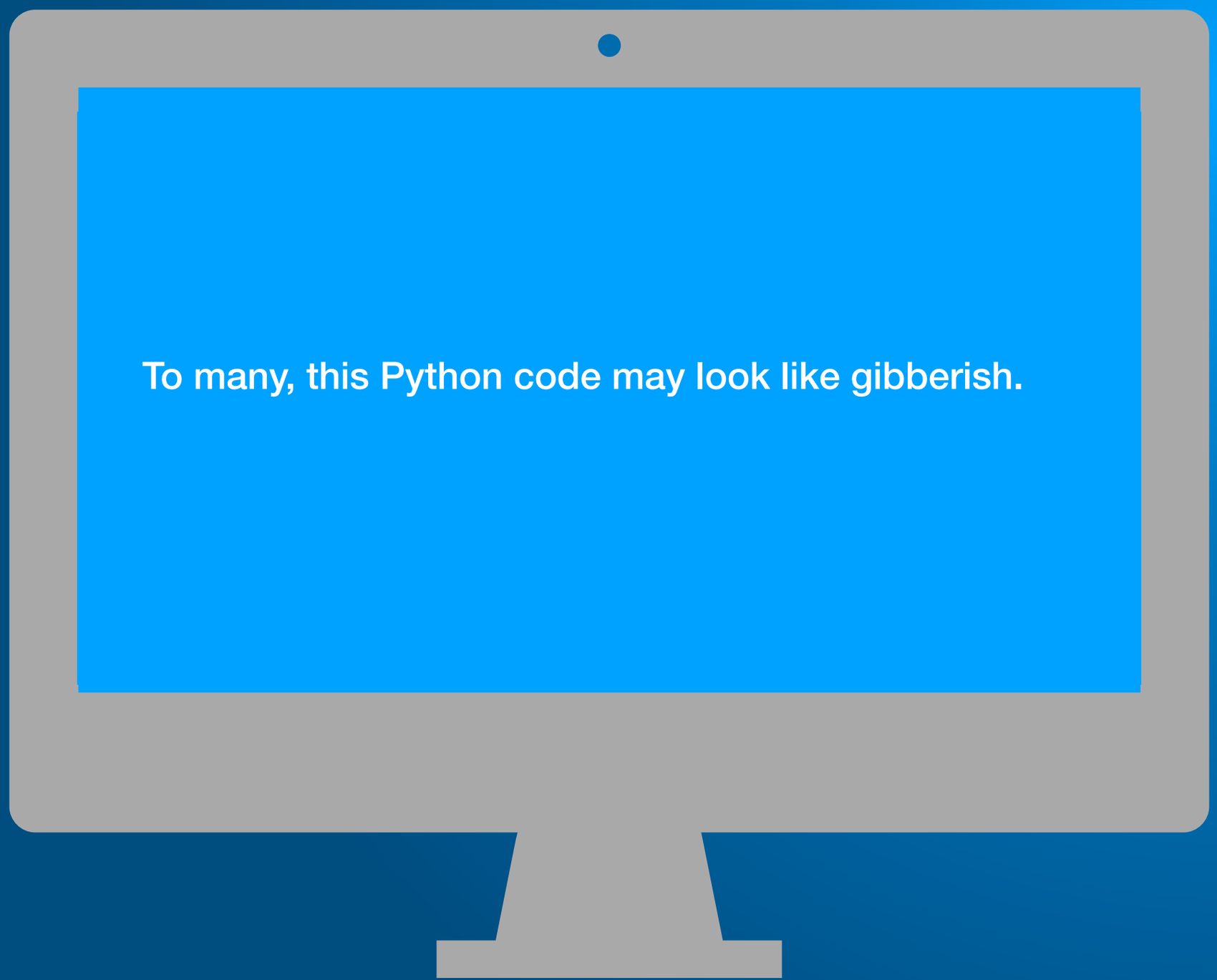
Data Liquidity

So what might liquidity look like?

```
import urllib
import json
import time
from twilio.rest import TwilioRestClient

# Twilio info
account_sid = "ACda2817bd10290bff2a66021d6204c94a"
auth_token  = "<SECRET DATA>"
from_phone = "+13107517490"
to_phone = "+18182937524"
client = TwilioRestClient(account_sid, auth_token)

# How far back we want to look for results
start_time = str(time.time() - 60 * 60) # An hour ago

# Practioner to fetch labs for
practioner_id = 259386

# URL to fetch lab results from
url="".join(["https://api.liquidity.duhs.com?_format=json&preformer=",
practioner_id, "&issued=>", start_time])
results = json.loads(urllib.urlopen(url));

# Loop through all results in the past hour to see if they are normal or not
for result in results:
  if data["interpretation"]["coding"][0]["code"] != "N":
```

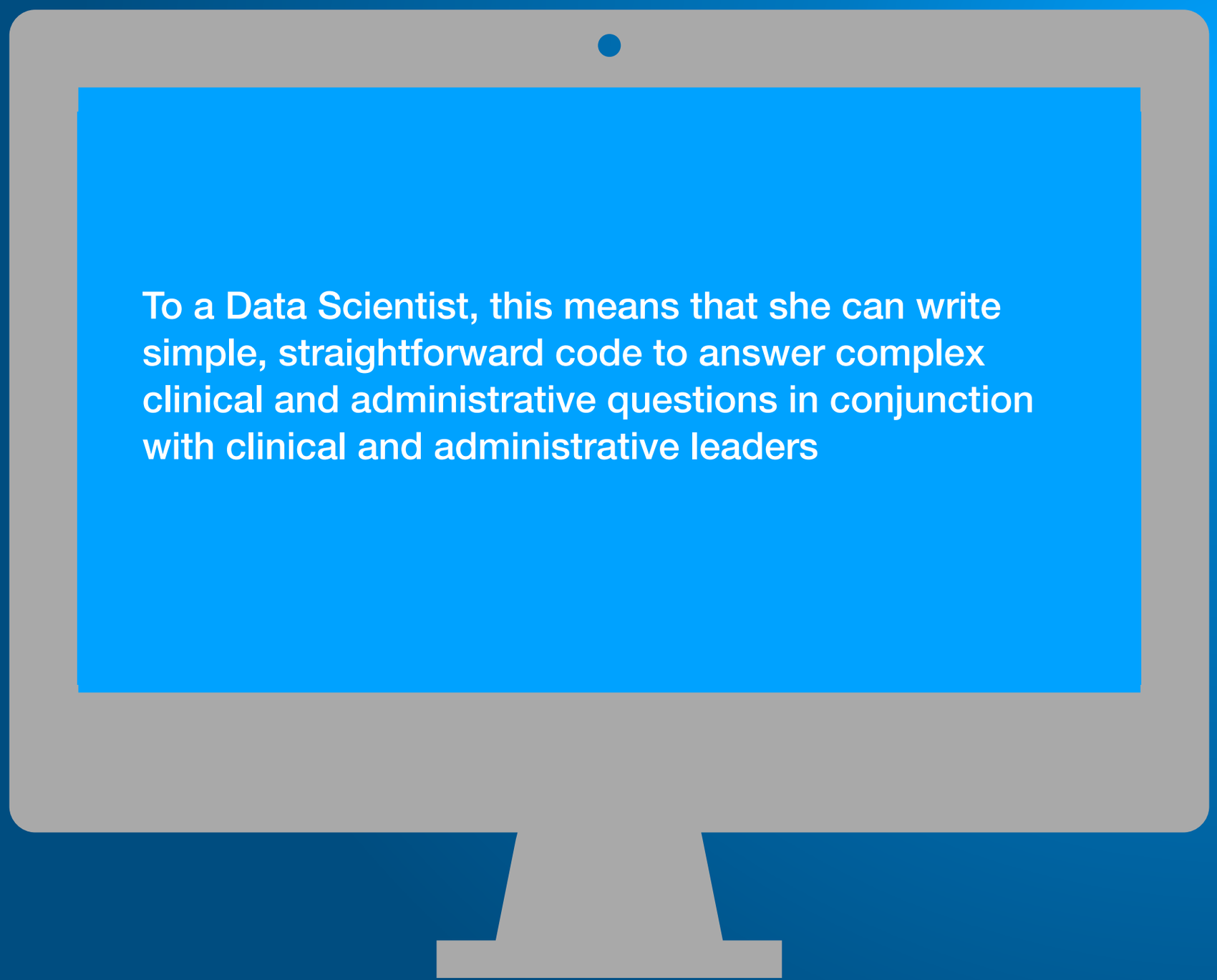
Data Liquidity

So what might liquidity look like?



Data Liquidity

So what might liquidity look like?



Data Liquidity

To get there, we need to consider two components

Data Liquidity

To get there, we need to consider two components

Policy

Data Liquidity

To get there, we need to consider two components

Policy

Technology

Data Liquidity

before delving into those, it's probably helpful to discuss what "data liquidity" is ***<u>not</u>***

Data Liquidity is <u>not</u>

<u>It is not</u> *untrammeled access to data for anyone regardless of its sensitivity*

Data Liquidity is <u>not</u>

<u>It is not</u> *untrammeled access to data for anyone regardless of its sensitivity*

<u>It is not</u> *solved only with technology*

Data Liquidity is <u>not</u>

<u>It is not</u> *untrammeled access to data for anyone regardless of its sensitivity*

<u>It is not</u> *solved only with technology*

<u>It is not</u> *a pipe dream*

Data Liquidity is

Is *agile & appropriate movement, merging, and analysis of data*

Data Liquidity is

Is *agile & appropriate movement, merging, and analysis of data*

Is *where infrastructure & access are not the rate-limiting step*

Data Liquidity is

Is *agile & appropriate movement, merging, and analysis of data*

Is *where infrastructure & access are not the rate-limiting step*

Is *where analytic priorities, not process, drive use*

Data Liquidity is

Is *agile & appropriate movement, merging, and analysis of data*

Is *where infrastructure & access are not the rate-limiting step*

Is *where analytic priorities, not process, drive use*

Is *secure, compliant, and auditable*

Data Liquidity

So let's consider these two components to data liquidity…

Data Liquidity

So let's consider these two components to data liquidity…

Policy

Data Liquidity

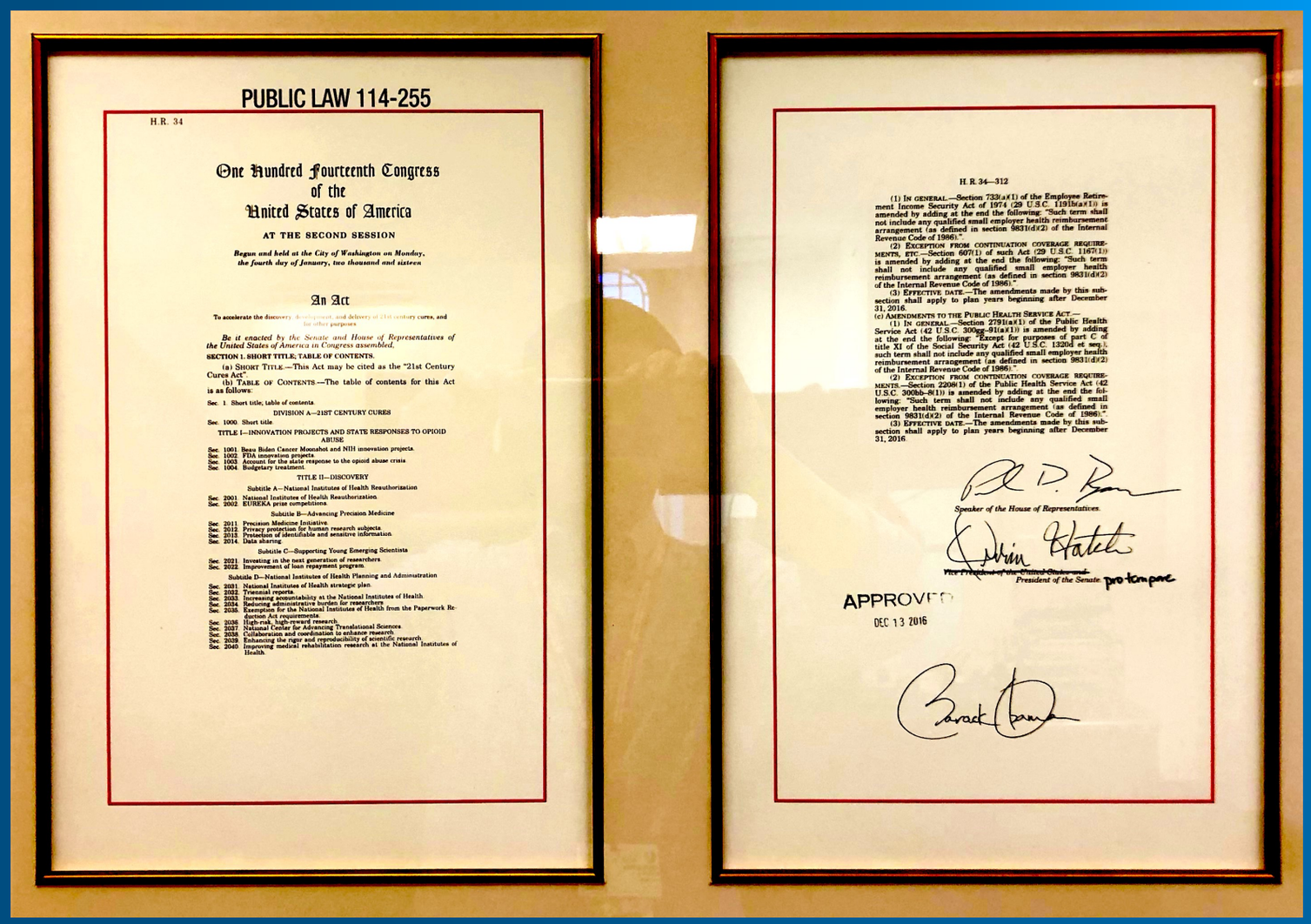
So let's consider these two components to data liquidity…

Policy

Technology

Data Liquidity

Policy



PUBLIC LAW 114-255

H.R. 34

One Hundred Fourteenth Congress
of the
United States of America

AT THE SECOND SESSION

Begun and held at the City of Washington on Monday, the fourth day of January, two thousand and sixteen

An Act

To accelerate the discovery, development, and delivery of 21st century cures, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "21st Century Cures Act".

(b) TABLE OF CONTENTS.—The table of contents for this Act is as follows:

H. R. 34—312

(1) IN GENERAL.—Section 733(a)(1) of the Employee Retirement Income Security Act of 1974 (29 U.S.C. 1191b(a)(1)) is amended by adding at the end the following: "Such term shall not include any qualified small employer health reimbursement arrangement (as defined in section 9831(d)(2) of the Internal Revenue Code of 1986)."

(2) EXCEPTION FROM CONTINUATION COVERAGE REQUIREMENTS, ETC.—Section 607(1) of such Act (29 U.S.C. 1167(1)) is amended by adding at the end the following: "Such term shall not include any qualified small employer health reimbursement arrangement (as defined in section 9831(d)(2) of the Internal Revenue Code of 1986)."

(3) EFFECTIVE DATE.—The amendments made by this subsection shall apply to plan years beginning after December 31, 2016.

(c) AMENDMENTS TO THE PUBLIC HEALTH SERVICE ACT.—

(1) IN GENERAL.—Section 2791(a)(1) of the Public Health Service Act (42 U.S.C. 300gg-91(a)(1)) is amended by adding at the end the following: "Except for purposes of part C of title XI of the Social Security Act (42 U.S.C. 1320d et seq.), such term shall not include any qualified small employer health reimbursement arrangement (as defined in section 9831(d)(2) of the Internal Revenue Code of 1986)."

(2) EXCEPTION FROM CONTINUATION COVERAGE REQUIREMENTS.—Section 2208(1) of the Public Health Service Act (42 U.S.C. 300bb-8(1)) is amended by adding at the end the following: "Such term shall not include any qualified small employer health reimbursement arrangement (as defined in section 9831(d)(2) of the Internal Revenue Code of 1986)."

(3) EFFECTIVE DATE.—The amendments made by this subsection shall apply to plan years beginning after December 31, 2016.

Speaker of the House of Representatives.

~~Vice President of the United States and~~
President of the Senate. pro tempore

APPROVED
DEC 13 2016

Data Liquidity

Policy

Sections | Browse | Search | Reader Aids | My FR | Search Documents

0 Sign in Sign up

NATIONAL ARCHIVES

FEDERAL REGISTER

The Daily Journal of the United States Government

Rule

21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program

A Rule by the Health and Human Services Department on 05/01/2020

PUBLISHED DOCUMENT

Start Printed Page 25642

AGENCY:

Office of the National Coordinator for Health Information Technology (ONC), Department of Health and Human Services (HHS).

ACTION:

Final rule.

SUMMARY:

This final rule implements certain provisions of the 21st Century Cures Act, including Conditions and Maintenance of Certification requirements for health information technology (health IT) developers under the ONC Health IT Certification Program (Program), the voluntary certification of health IT for use by pediatric health care providers, and reasonable and necessary activities that do not constitute information blocking. The implementation of these provisions will advance interoperability and support the access, exchange, and use of electronic health information. The rule also finalizes certain modifications to the 2015 Edition health IT certification criteria and Program in additional ways to

DOCUMENT DETAILS

Printed version:
PDF

Publication Date:
05/01/2020

Agencies:
Department of Health and Human Services
Office of the Secretary

Dates:
Effective date: This final rule is effective on June 30, 2020.

Effective Date:
06/30/2020

Document Type:
Rule

Document Citation:
85 FR 25642

Page:
25642-25961 (320 pages)

CFR:
45 CFR 170

Policy

Data Liquidity

Policy

21st Century Cures Act, Section 4002

"... must also attest that it published application program interfaces ( ) and allows health information from such APIs to be accessible, exchanged and used without special effort through the use of APIs or successor technologies or standards, including providing access to all data elements of a patient's EHR to the extent permissible under applicable privacy laws."

Data Liquidity

Let’s now look at

Technology

Data Liquidity

Technology

Data Standards

The diversity of data types: e.g. structured, unstructured, EHR, molecular, wearable, social determinants, &c…

Data Liquidity

Technology

Data Standards

The diversity of data types: e.g. structured, unstructured, EHR, molecular, wearable, social determinants, &c…

Formats not only have to account for structure, but transmissibility. Systems have to be prepared for realtime data transactions

Data Liquidity

Technology

Data Standards

The diversity of data types: e.g. structured, unstructured, EHR, molecular, wearable, social determinants, &c…

Formats not only have to account for structure, but transmissibility. Systems have to be prepared for realtime data transactions

Fungibility: machine learning learns with bulk data, but must be able to generate inference with individual data

Data Liquidity

Technology

Considerations include:

The diversity of data t
structured, unstructu
molecular, wearable,
determinants, &c…

The velocity with whic
originate and move. S
prepared for realtime
analysis

The scalability and fle
compute infrastructu
diverse data science
these diverse data typ

HL7 FHIR

Data Liquidity

Technology

Considerations include:

HL7 FHIR

Not a panacea!

The diversity of data t
structured, unstructu
molecular, wearable,
determinants, &c…

The velocity with whi
originate and move. S
prepared for realtime
analysis

The scalability and fle
compute infrastructu
diverse data science
these diverse data typ

Data Liquidity

Technology

Considerations include:

HL7 FHIR

- *Immature ecosystem*
- *More of a transaction standard*
- *What's "inside the box" can be quite permissive*
- *Needs real world critical mass for us to learn good vs bad implementations*

Data Liquidity

IT IS NOW TIME TO BUILD

Data Liquidity

We have a tremendous opportunity to build a true clinical learning ecosystem

IT IS NOW TIME TO BUILD

Data Liquidity

“the more that we use data, the clearer the river of data gets”

—Amy Abernethy

Future: Overcoming barriers to implementation

Ref. 20

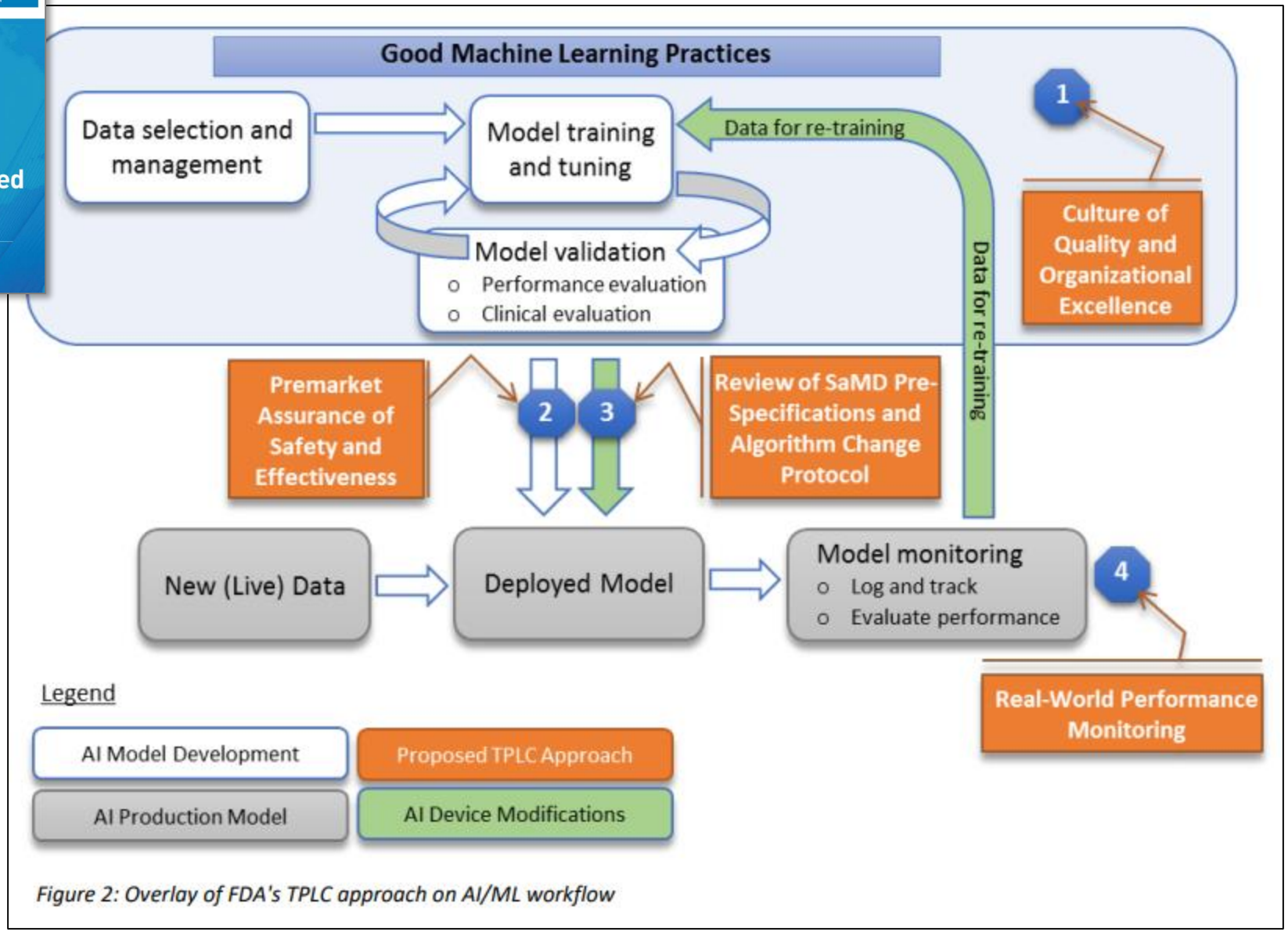
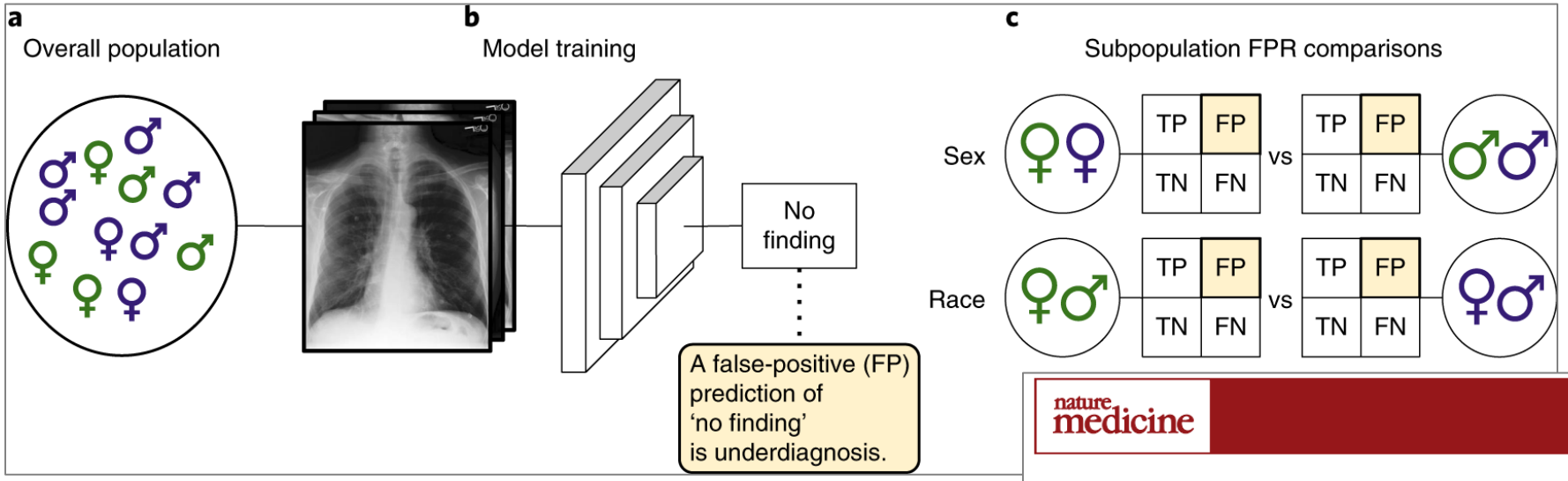
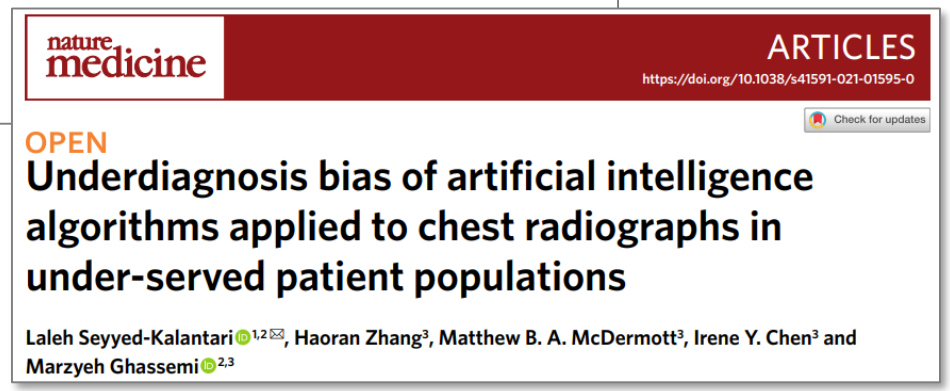


Figure 2: Overlay of FDA's TPLC approach on AI/ML workflow

Future: Overcoming barriers to implementation




nature medicine
ARTICLES
https://doi.org/10.1038/s41591-021-01595-0
Check for updates
OPEN
Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations
Laleh Seyyed-Kalantari[1,2], Haoran Zhang[3], Matthew B. A. McDermott[3], Irene Y. Chen[3] and Marzyeh Ghassemi[2,3]

Ref. 21

Future: Overcoming barriers to implementation

Man is to Computer Programmer as Woman is to Homemaker?
Debiasing Word Embeddings

Tolga Bolukbasi[1], Kai-Wei Chang[2], James Zou[2], Venkatesh Saligrama[1,2], Adam Kalai[2]
[1]Boston University, 8 Saint Mary's Street, Boston, MA
[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

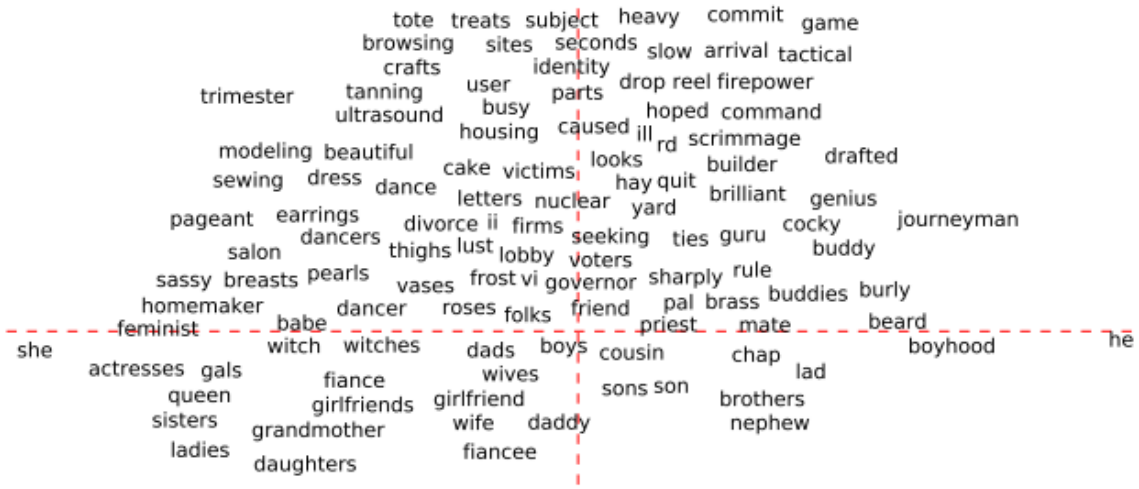


Figure 7: Selected words projected along two axes: x is a projection onto the difference between the embeddings of the words *he* and *she*, and y is a direction learned in the embedding that captures gender neutrality, with gender neutral words above the line and gender specific words below the line. Our hard debiasing algorithm removes the gender pair associations for gender neutral words. In this figure, the words above the horizontal line would all be collapsed to the vertical line.

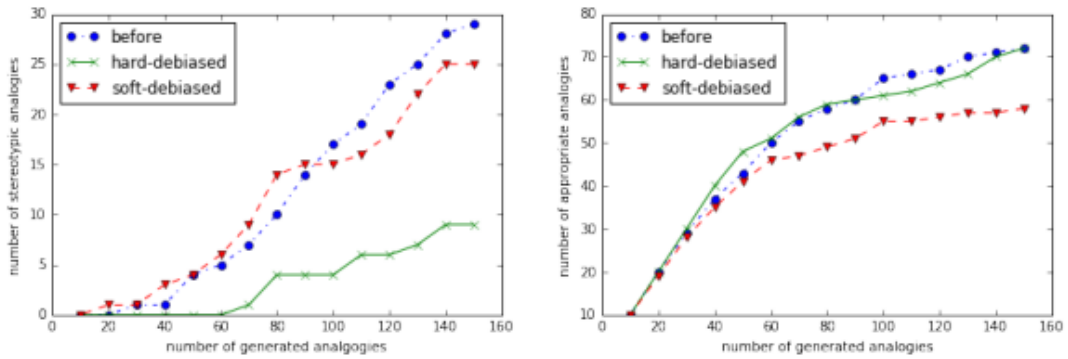
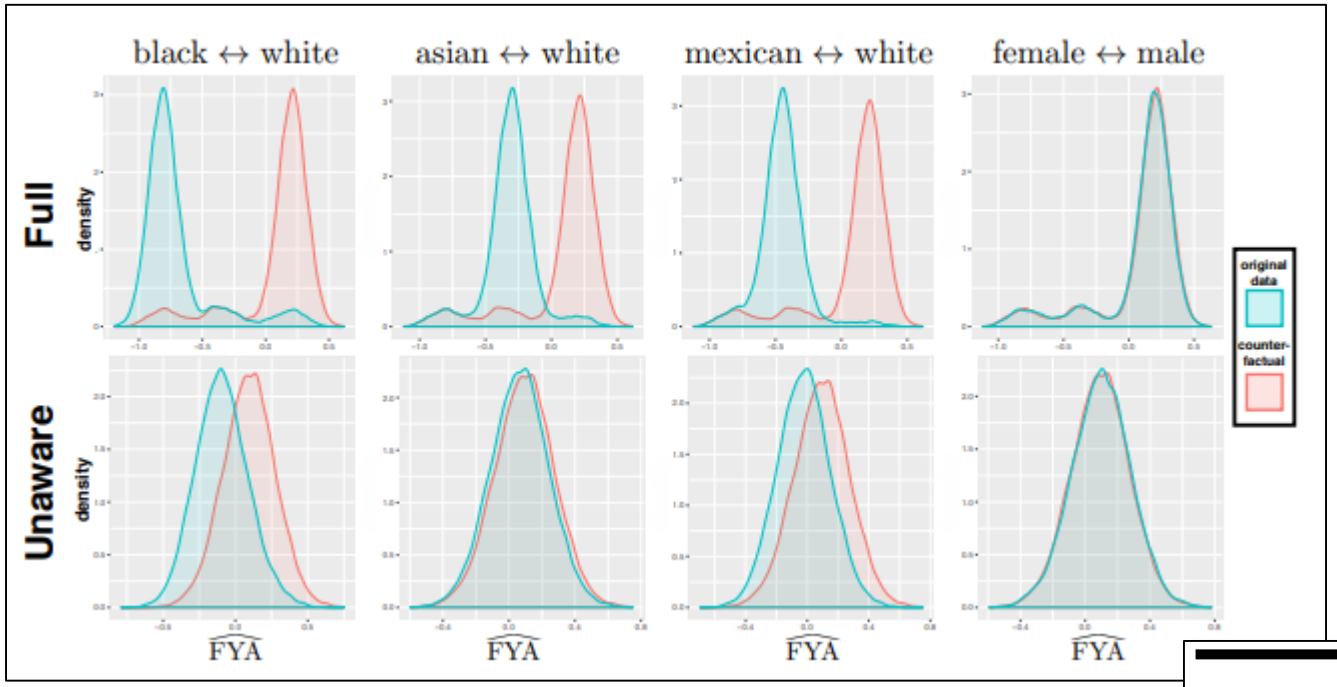
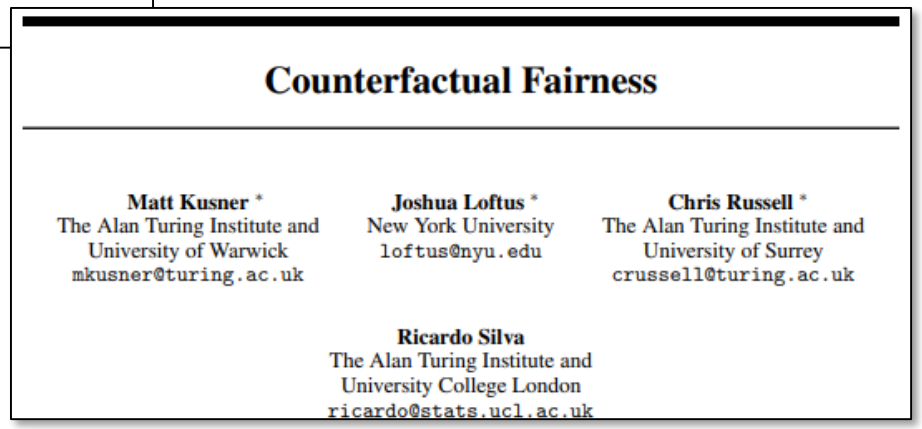


Figure 8: Number of stereotypical (Left) and appropriate (Right) analogies generated by wordembeddings before and after debiasing.

Ref. 22

Future: Overcoming barriers to implementation





Counterfactual Fairness

Matt Kusner *
The Alan Turing Institute and
University of Warwick
mkusner@turing.ac.uk

Joshua Loftus *
New York University
loftus@nyu.edu

Chris Russell *
The Alan Turing Institute and
University of Surrey
crussell@turing.ac.uk

Ricardo Silva
The Alan Turing Institute and
University College London
ricardo@stats.ucl.ac.uk

Ref. 23

Future: Overcoming barriers to implementation

FAT/ML

airness ccountability ransparency

Future: Immediate next steps

- **Educate yourself and trainees about ML techniques & reporting standards.**
- **Engage with efforts to define the regulatory perspective on ML in CR.**
- **Collaborate on proof-of-concept studies showing the promise of ML in CR & comparing ML to conventional approaches.**
- **Support data interoperability initiatives and advocate for patient-centered approaches to data ownership.**

References

1. Weissler EH, Naumann T, Andersson T, et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials. Aug 16 2021;22(1):537. doi:10.1186/s13063-021-05489-x*
2. Annapureddy AR, Angraal S, Caraballo C, et al. The National Institutes of Health funding for clinical research applying machine learning techniques in 2017. *NPJ Digit Med. 2020;3:13. doi:10.1038/s41746-020-0223-9*
3. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for Reporting Machine Learning Analyses in Clinical Research. *Circ Cardiovasc Qual Outcomes. Oct 2020;13(10):e006556. doi:10.1161/CIRCOUTCOMES.120.006556*
4. Bisaso KR, Anguzu GT, Karungi SA, Kiragga A, Castelnuovo B. A survey of machine learning applications in HIV clinical research and care. *Comput Biol Med. Dec 1 2017;91:366-371. doi:10.1016/j.compbiomed.2017.11.001*
5. Kim KJ, Tagkopoulos I. Application of machine learning in rheumatic disease research. *Korean J Intern Med. Jul 2019;34(4):708-722. doi:10.3904/kjim.2018.349*
6. *Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. Apr 1 2019;20(2):273-286. doi:10.1093/biostatistics/kxx069*
7. Fauqueur J TA, Togia T. Constructing large scale biomedical knowledge bases from scratch with rapid annotation of interpretable patterns. *Proceedings of the 18th BioNLP Workshop and Shared Task. 2019;doi:doi:10.18653/v1/w19-5016.*
8. Madhukar NS, Khade PK, Huang L, et al. A Bayesian machine learning approach for drug target identification using diverse data types. *Nat Commun. Nov 19 2019;10(1):5221. doi:10.1038/s41467-019-12928-6*
9. Liu Q AM, Brockschmidt M, Gaunt AL. Constrained graph variational autoencoders for molecule design. *NeurIPS 2018. 2018;doi:arXiv:1805.09076*
10. Langner S, Hase F, Perea JD, et al. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv Mater. Apr 2020;32(14):e1907801. doi:10.1002/adma.201907801*
11. Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics. Dec 2011;67(4):1422-33. doi:10.1111/j.1541-0420.2011.01572.x*
12. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA. May 28 2019;321(20):2003-2017. doi:10.1001/jama.2019.5791*
13. Glicksberg BS, Miotto R, Johnson KW, et al. Automated disease cohort selection using word embeddings from Electronic Health Records. *Pac Symp Biocomput. 2018;23:145-156.*
14. Chen R, Jankovic F, Marinsek N, et al. Developing Measures of Cognitive Impairment in the Real World from Consumer-Grade Multimodal Sensor Streams. presented at: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining; 2019; Anchorage, AK, USA. https://doi.org/10.1145/3292500.3330690
15. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ. May 8 2015;350:h2147. doi:10.1136/bmj.h2147*
16. Schork NJ. Personalized medicine: Time for one-person trials. *Nature. Apr 30 2015;520(7549):609-11. doi:10.1038/520609a*
17. Mitchell M, Wu S, Zaldivar A, et al. Model Cards for Model Reporting. presented at: Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019; Atlanta, GA, USA. https://doi.org/10.1145/3287560.3287596
18. Sterckx S, Rakic V, Cockbain J, Borry P. "You hoped we would sleep walk into accepting the collection of our data": controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. *Med Health Care Philos. Jun 2016;19(2):177-90. doi:10.1007/s11019-015-9661-6*
19. Robbins RB, E. An invisible hand: Patients aren't being told about the AI systems advising their care. *STAT.* 15 July 2020. https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/
20. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan (2021). https://www.fda.gov/media/145022/download
21. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med. Dec 2021;27(12):2176-2182. doi:10.1038/s41591-021-01595-0*
22. Bolukbasi T, Chang K-W, Zou J, Saligrama V, Kalai A. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. presented at: Proceedings of the 30th International Conference on Neural Information Processing Systems; 2016; Barcelona, Spain.
23. Kusner M LJ, Russell C, Silva R. Counterfactual fairness. *Adv Neural Inf Process Syst. 2017;doi:arXiv:1703.06856*

Thank you

Questions and comments?