ATHENA

Causal AI to increase clinical trial success

Manjari Narayan, PhD

Rapidly improve **quality of decision tools** in drug discovery and development using AI

↑ **10x** clinical trial **success rates**

↓ Save patients **billions of dollars**

Experienced quantitative scientist in biomedicine

Manjari Narayan, PhD

ILLINOIS
Electrical & Computer Engineering

RICE ENGINEERING
Electrical and
Computer Engineering

DYNO
THERAPEUTICS

RICE UNIVERSITY
DSP

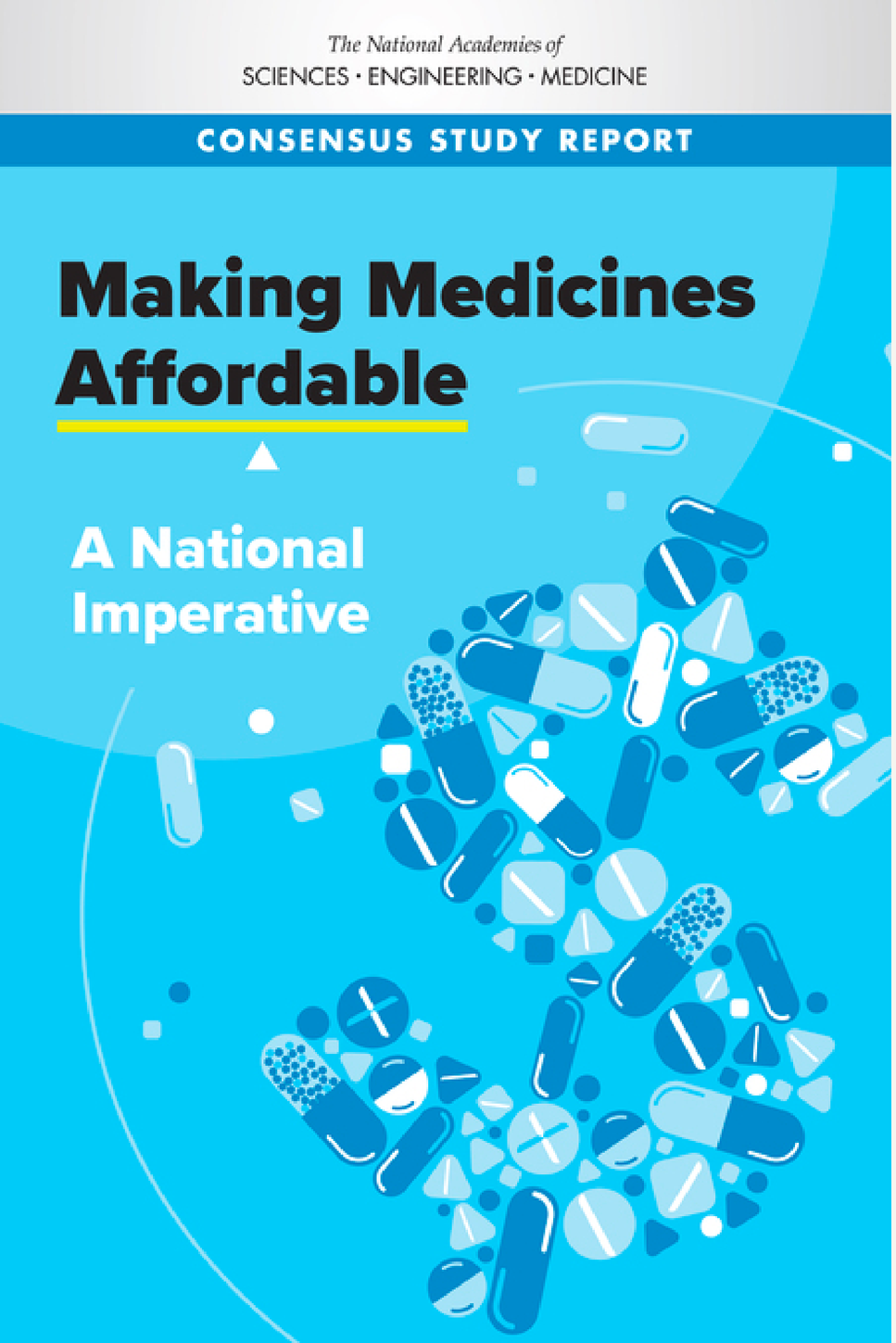
Stanford
MEDICINE

Signal Processing

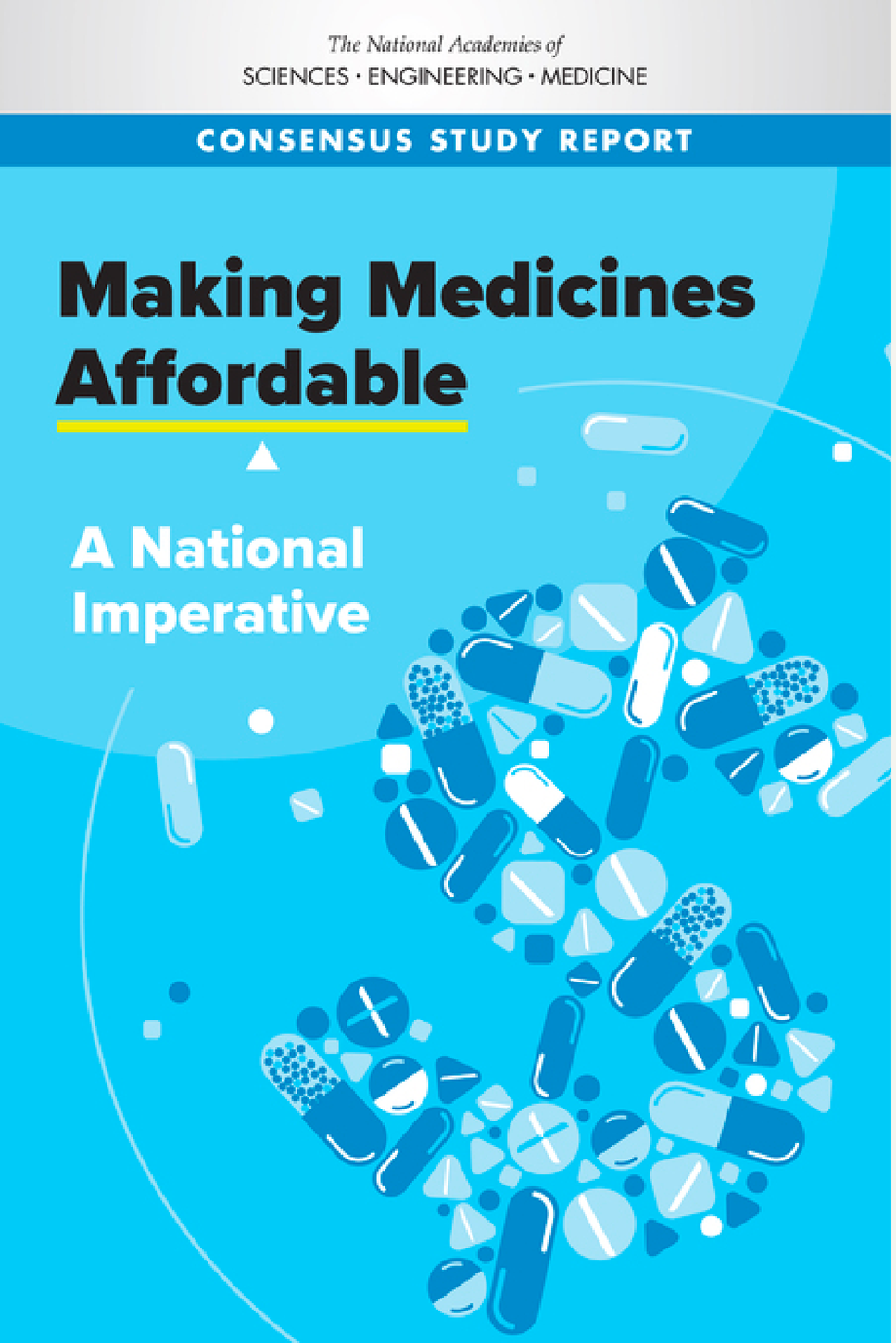
Statistical Machine Learning

Human Neuroscience; Psychiatry

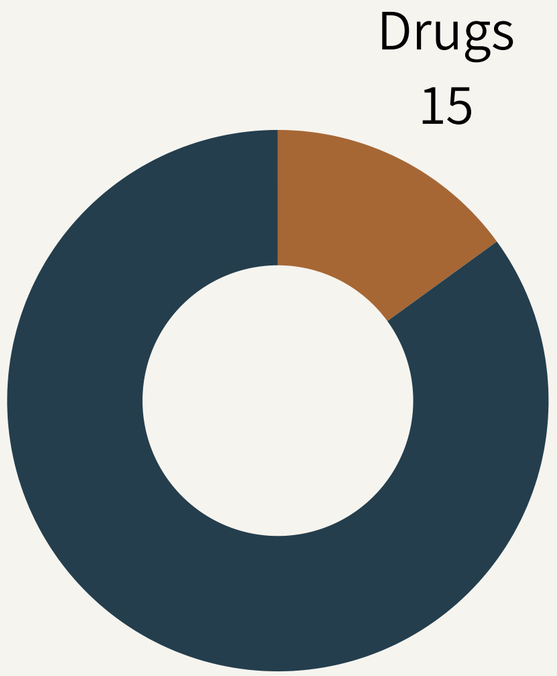
AI guided protein science; Gene Therapy



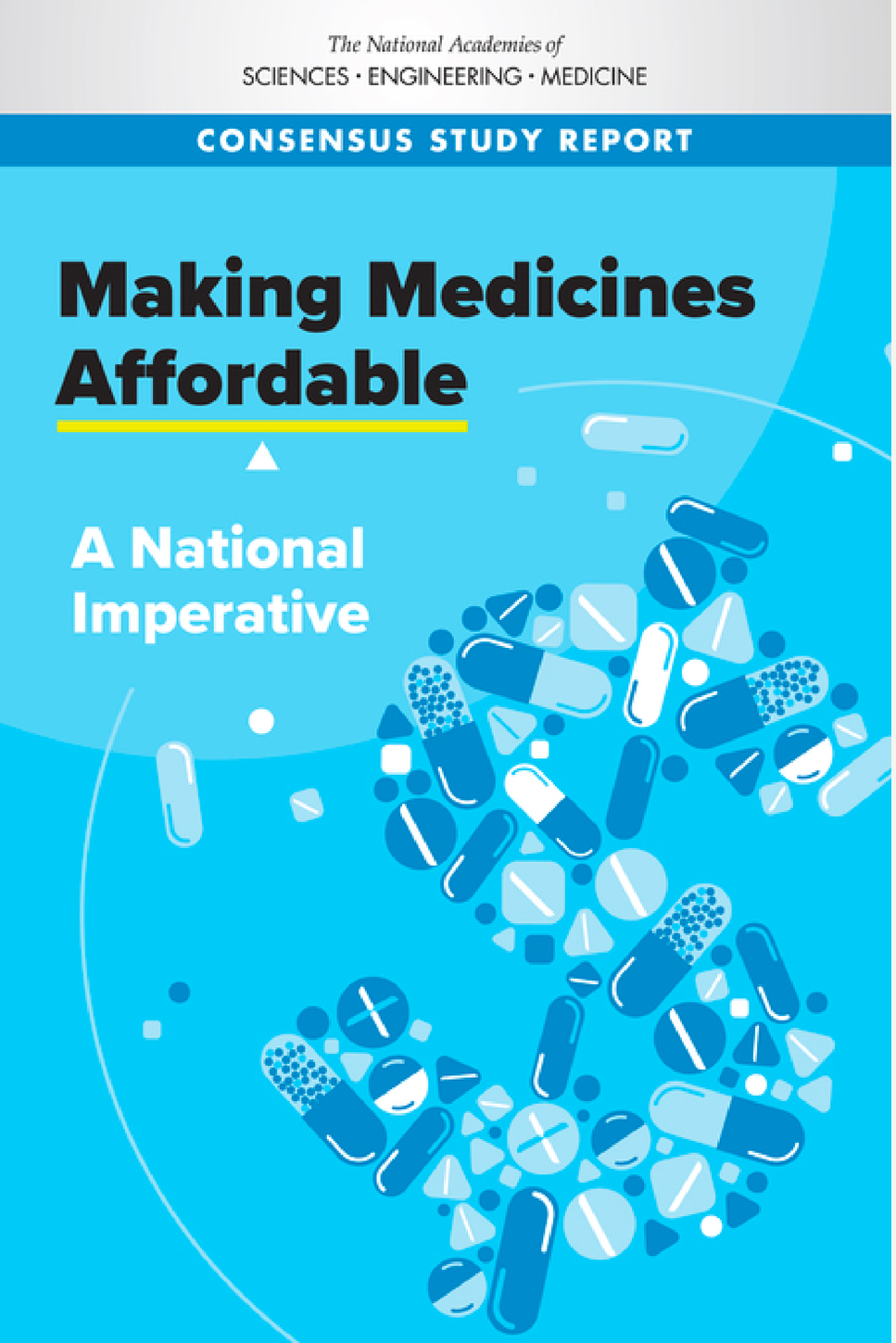
Healthcare expenditure — 20% of GDP by 2030



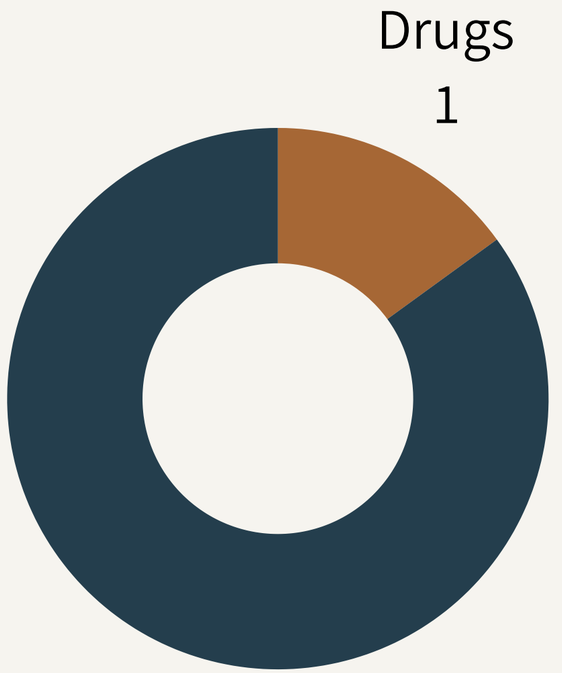
Healthcare expenditure — 20% of GDP by 2030



Source data: IQVIA Report, 2021



Healthcare expenditure — 20% of GDP by 2030

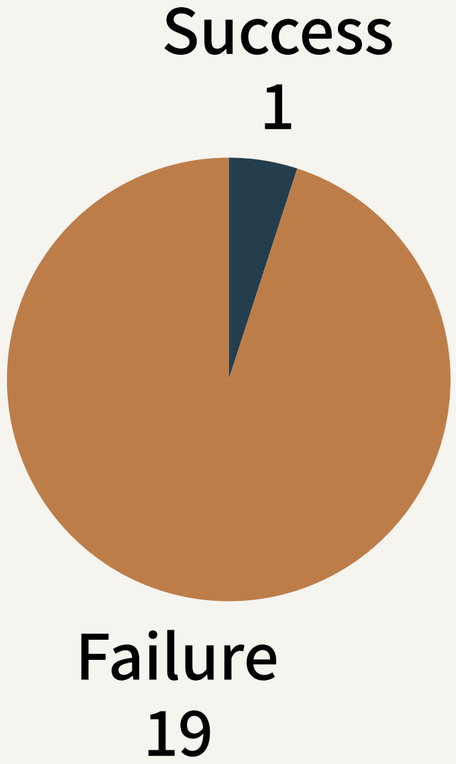


in trillions

Source data: KFF report, 2021

Success rates in drug development

5.7%



Source data: Wong et. al. (2019); Evaluate Pharma, March 2023

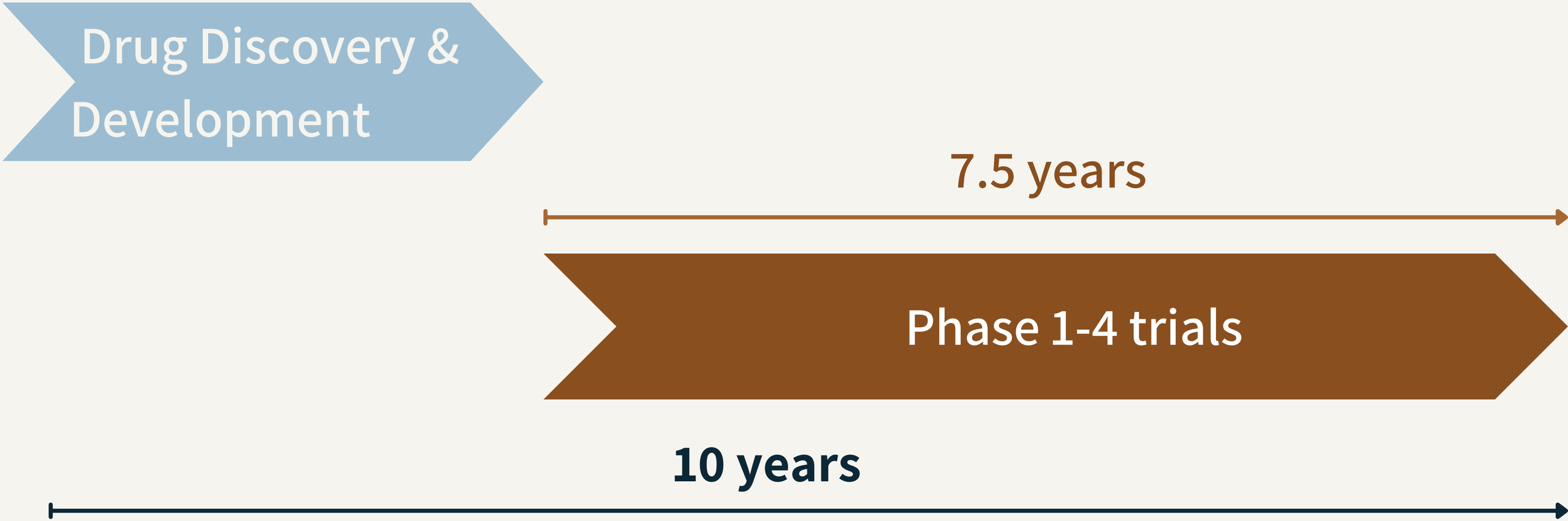
Success rates in drug development

5.7%

USD 2+ billion per new drug launch

Source data: Wong et. al. (2019); Evaluate Pharma, March 2023

Most of the time & money is spent in clinical trials



What if we could 10x success rates for every disease area?

57%
success rates

< 200m
per new
drug launch

Source data: Wong et. al. (2019); Evaluate Pharma, March 2023

ATHENA: Platform to enable rapid improvement of decision tools and increase successful clinical trials

Traditional drug development relies on a series of proxy signals

Target Discovery

Identify better target to modify disease

Traditional drug development relies on a series of proxy signals

Target Discovery

Identify better target to modify disease

Drug Discovery

Search and improve the best drug/molecule

Traditional drug development relies on a series of proxy signals

Target Discovery

Identify better target to modify disease

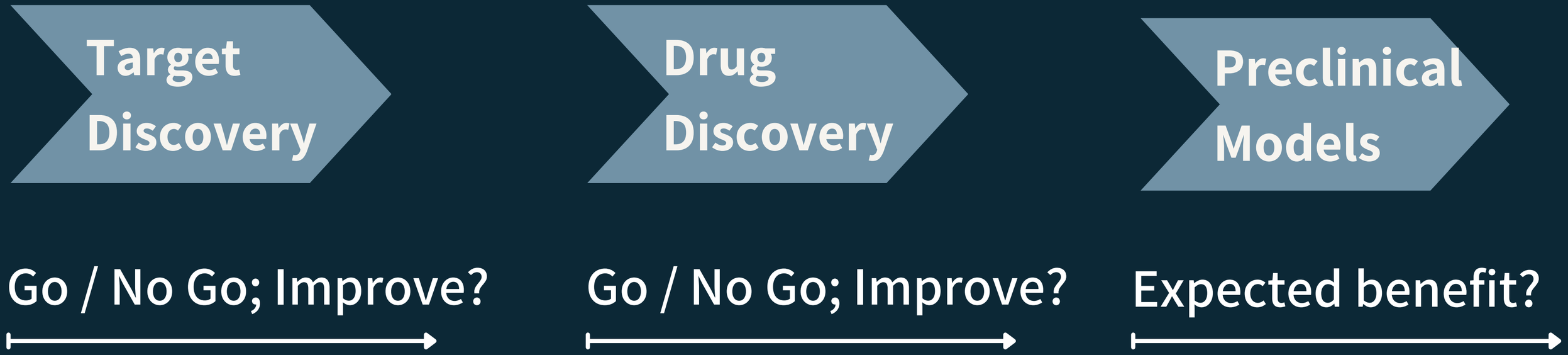
Drug Discovery

Search and improve the best drug/molecule

Preclinical Models

Does it work in mice?

Proxies are decision tools to guide early drug development



Examples of decision tools

Biomarkers of disease mechanism

Assays of drug properties
(e.g solubility)

3D models, organoids, organ-on-a-chip

Animal models of disease
(e.g mice)

Examples of decision tools

How good are these tools?

Biomarkers of disease mechanism

Assays of drug properties
(e.g solubility)

3D models, organoids, organ-on-a-chip

Animal models of disease
(e.g mice)

Between 1999–2009, **16** IGF-1 inhibitors tested in **183** cancer trials all failed despite **“favorable” preclinical data**

Jentzsch V, Osipenko L, Scannell JW, Hickman JA. Costs and Causes of Oncology Drug Attrition With the Example of Insulin-Like Growth Factor-1 Receptor Inhibitors. JAMA Netw Open. 2023. doi:10.1001/jamanetworkopen.2023.24977

Scannell JW, Bosley J, Hickman JA, et al. Predictive validity in drug discovery: what it is, why it matters and how to improve it. Nat Rev Drug Discov. 2022. doi:10.1038/s41573-022-00552-x

How do people assess tools today?

How do people assess tools today?

1. Intuition

How do people assess tools today?

1. Intuition
2. Reliability

How do people assess tools today?

1. Intuition
2. Reliability
3. Biased metrics — non-causal

How do people assess tools today?

1. Intuition
2. Reliability
3. Biased metrics — non-causal
4. Clinical success/failures — coarse

Not a good fit for academia or biotech or pharma

Academia

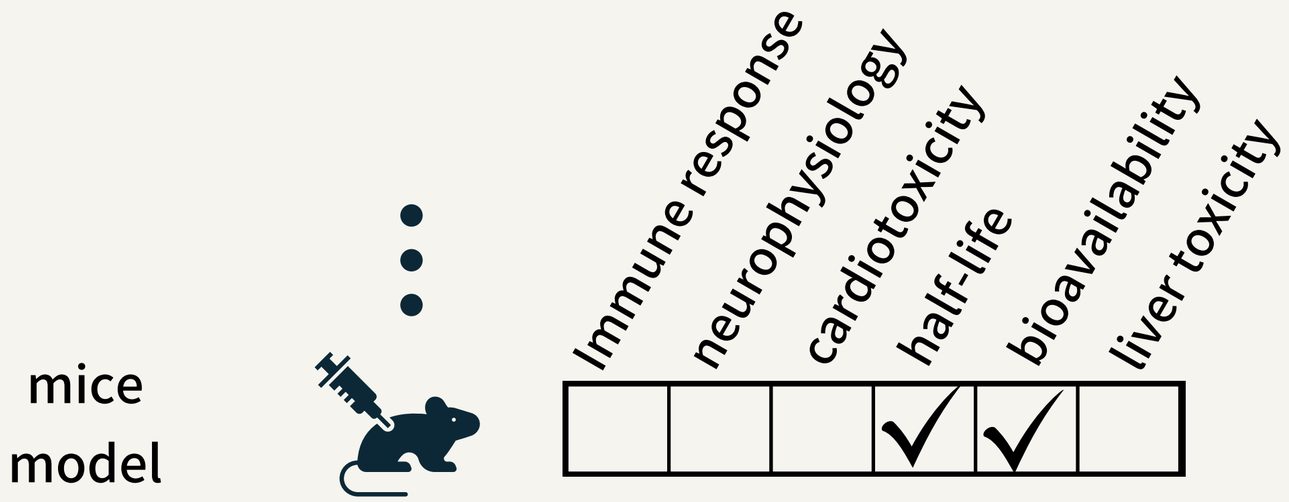
Typically don’t test discoveries with clinical development

Biopharma

Evaluate a potential improvement only through 5-10 year cycle with new clinical trials

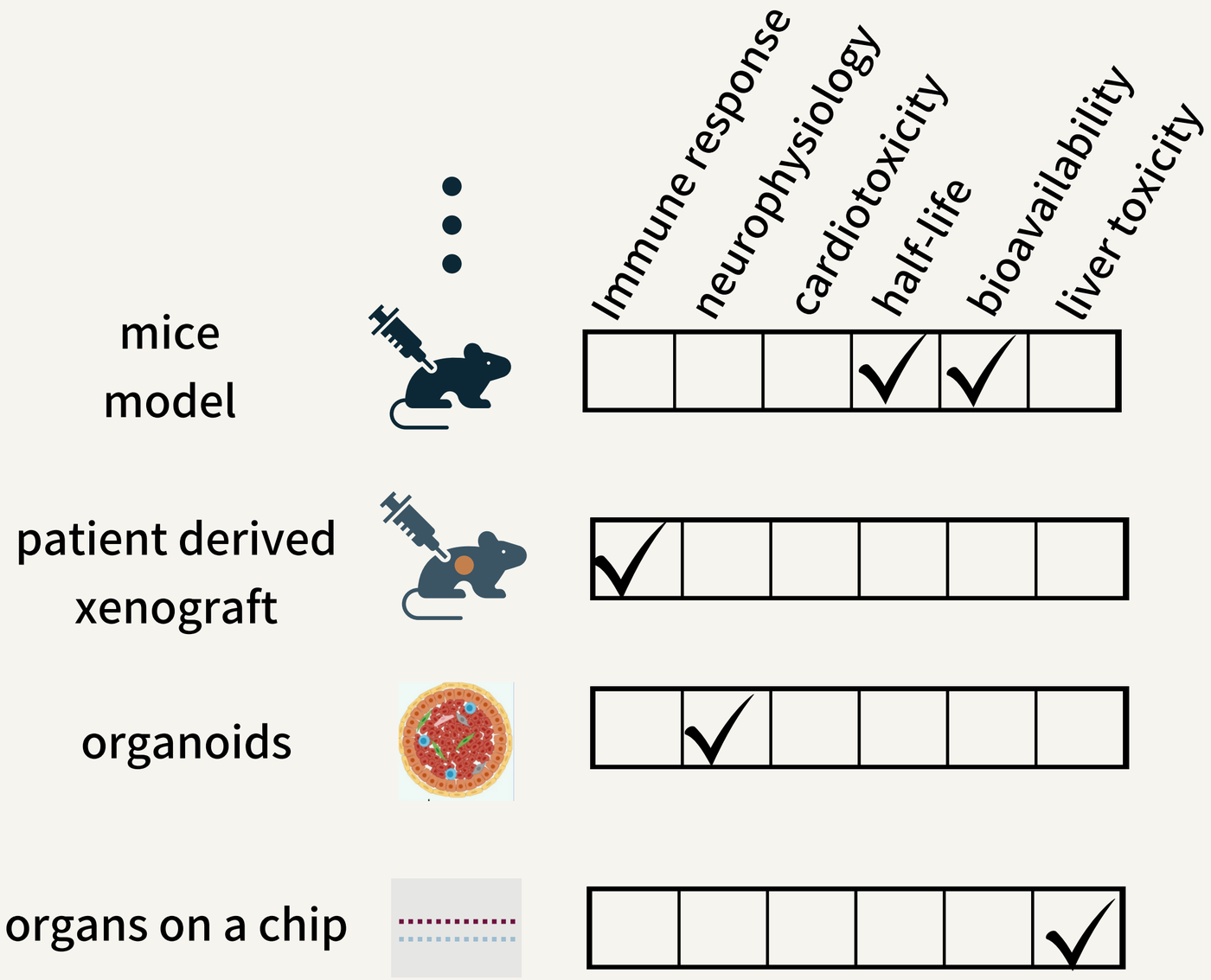
ATHENA.1: Systematically evaluate decision tools using clinical data

f(therapy, disease)



At best, any individual assay or tool maybe partially useful

f(therapy, disease)



At best, any individual assay or tool maybe partially useful

“I think the best way to make progress on applications of machine learning to drug discovery is to **fund a large public effort that will generate high-quality data and make this data available to the community.**

Patrick Walters,
Chief Data Officer of Relay Therapeutics

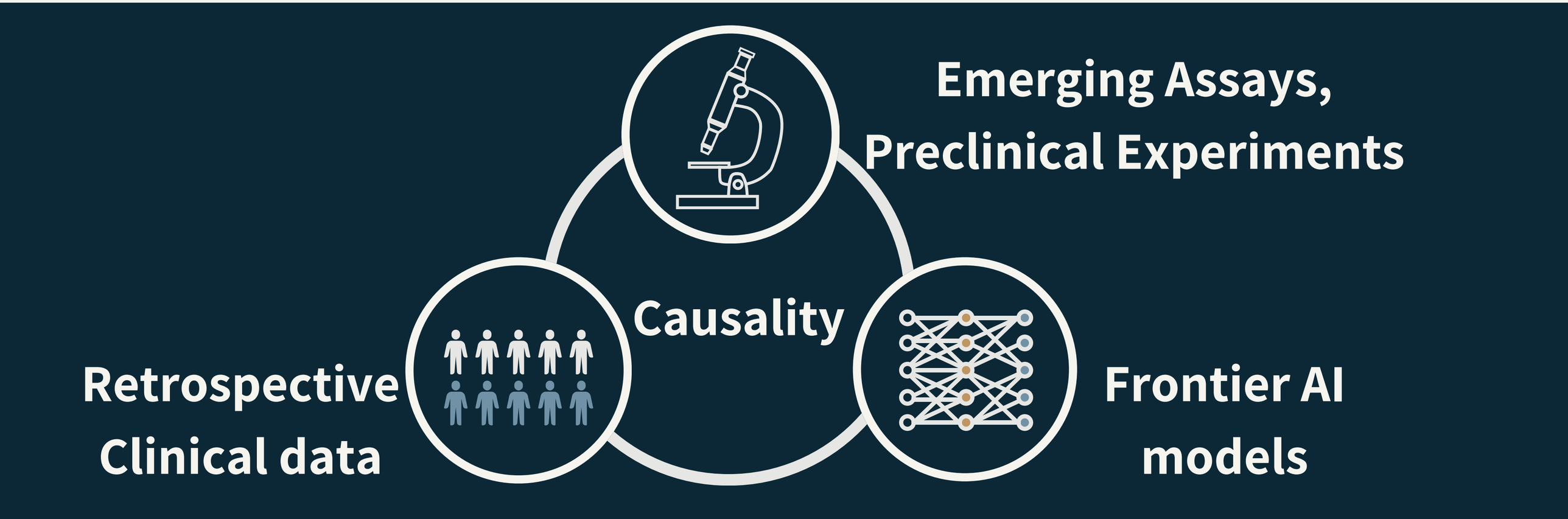
“Generally too few data to expect algorithms to infer relationships between drug treatment and in vivo endpoint observations

Andreas Bender,
Professor of Molecular Informatics, Cambridge

“We lack consistent and quality data for both training and validation across most applications in drug discovery.

Sam Bjork & Travis Hughes
Digitalis Ventures

ATHENA.2: Create an improved integrated decision tool to predict in-human data using Causal AI

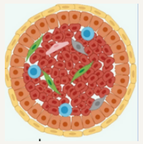


Learn optimal synthesis of preclinical models to predict > 90% human equivalent response to novel therapeutics

mice model

patient derived xenograft

organoids



organs on a chip

drug metabolism

immune response

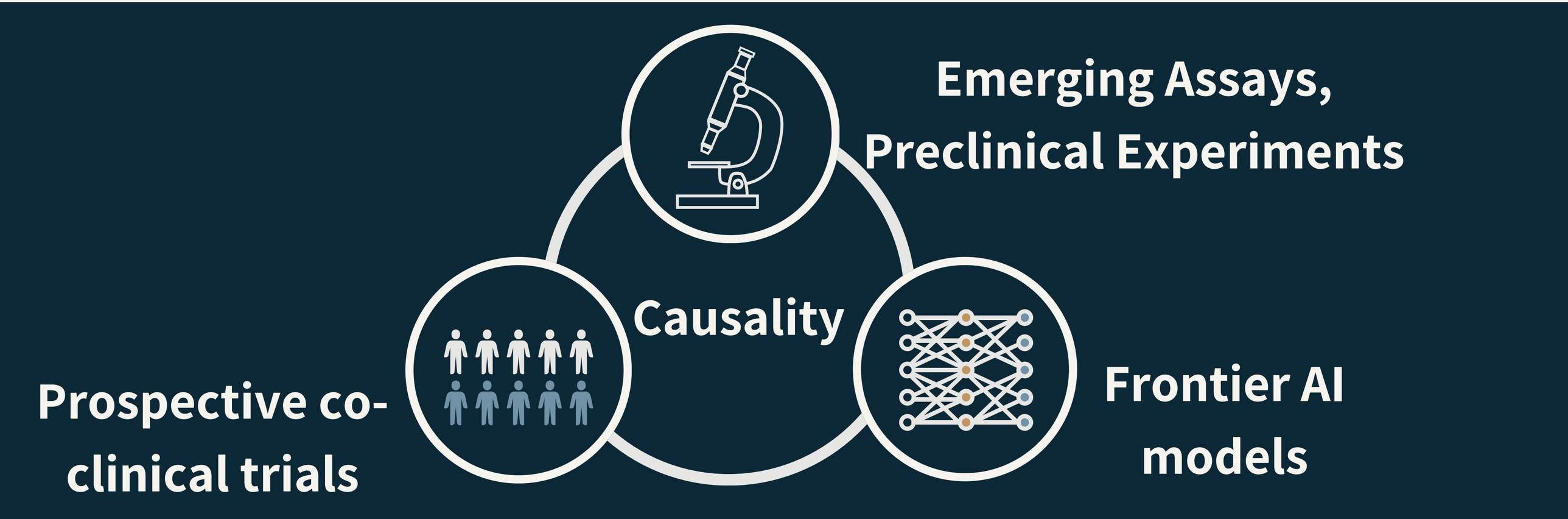
neuron electrophysiology

liver toxicity

clinical data

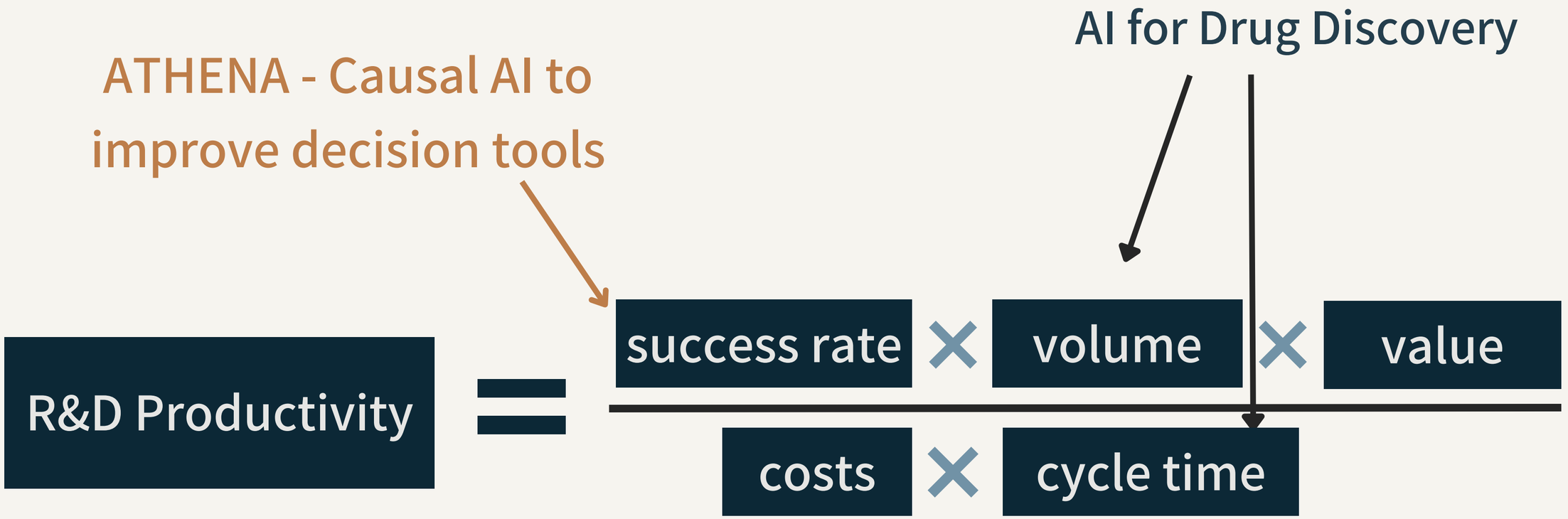
Causally aware AI to triangulate strengths/weaknesses

ATHENA.3: Use co-clinical trials to assess ATHENA vs. existing approaches



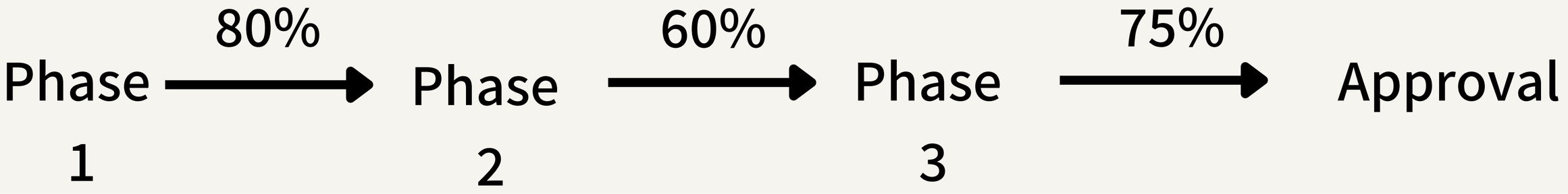
ATHENA vs. Other AI

ATHENA vs. Other AI

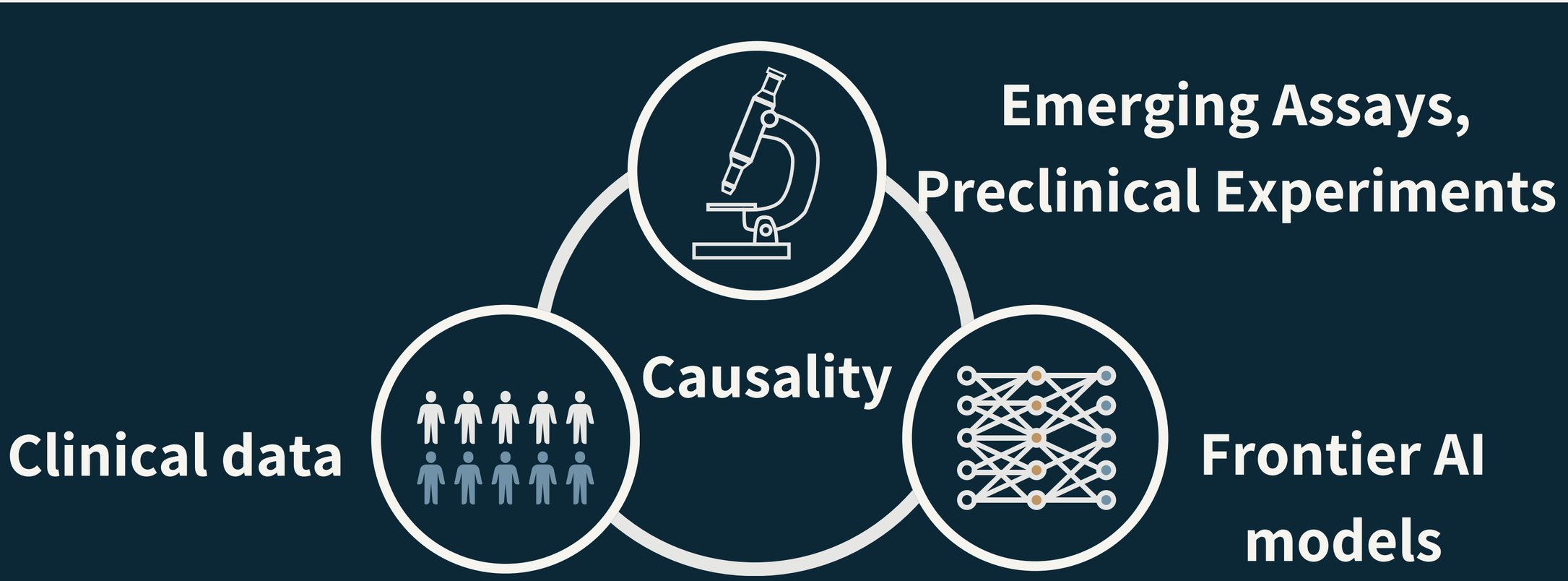


5x success rates in < 5 years

36% overall success rates in a single disease area



If ATHENA succeeds, tangible proof that decision tool quality matters!



Spur adoption in early drug development ecosystem

Biotech Startups

THYMMUNE

NewLimit

emulate

Biopharma & VCs

Roche

VERTEX

RELATED SCIENCES

Curie.Bio

DIGITALIS VENTURES

Non-profits

ALS THERAPY DEVELOPMENT INSTITUTE

ALS TDI Logo

Help Us Solve The Cruel Mystery LUPUS FOUNDATION OF AMERICA

BILL & MELINDA GATES foundation

Alzheimer's Drug Discovery Foundation

THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH

Translational / Regulatory Science

NIH National Center for Advancing Translational Sciences

FDA U.S. FOOD & DRUG ADMINISTRATION

Full vision needs 150m over 5 years

Phase 1: 30 m

Build & evaluate platform for a data-rich problem — oncology & toxicity

Phase 2: 60m per disease

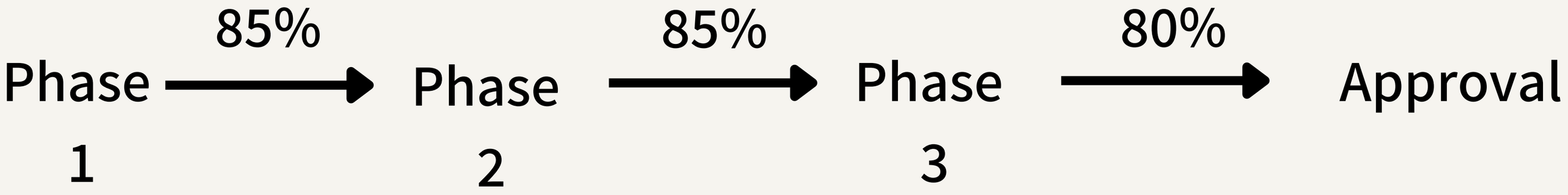
Scale to multiple decision tools in 2 disease areas — neurodegenerative & autoimmune

Improving decision tools —ARPA hard

- Decision tools are a public good: requires aligning incentives & public-private-government partnerships
- Create shared problem + trustworthy evaluation
- Bringing together experts in early drug development + experts in improving external validity from causal AI

10x success rates in 10 years

57% overall success rates across all diseases

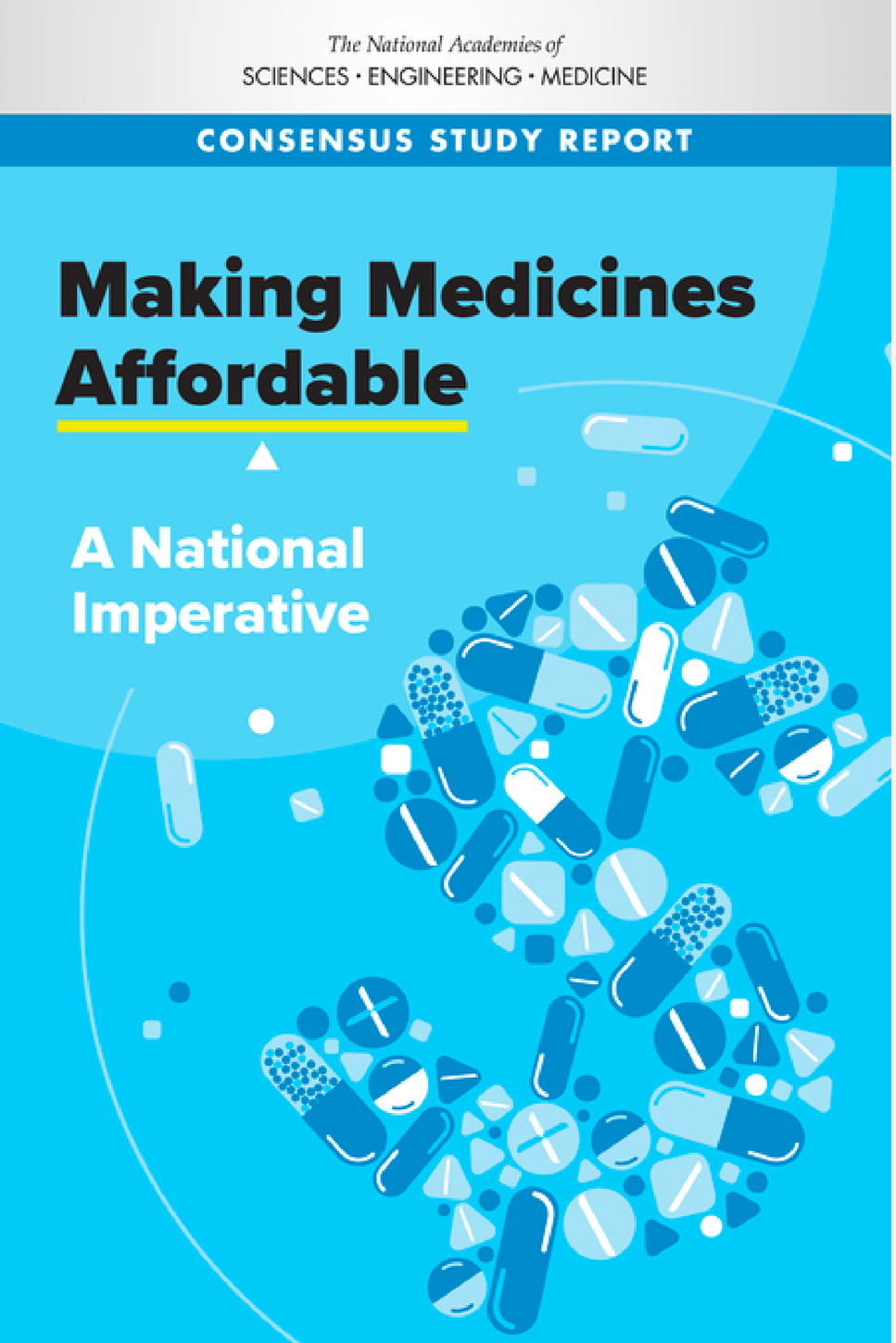


Thank you

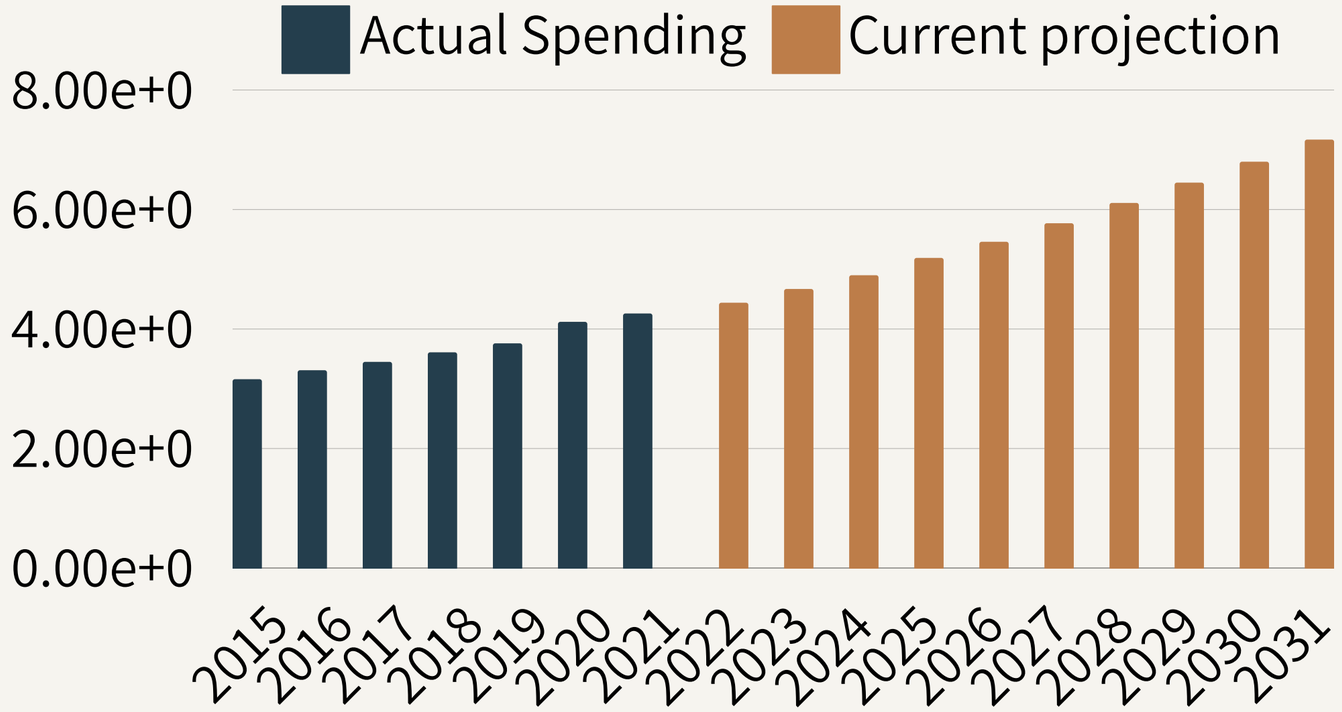
Manjari Narayan

manjari@manjarinarayan.com

Additional Slides



Healthcare expenditure projected to rise to 20% of GDP by 2030



Precedent: Why is this in the realm of feasible?

Related Efforts

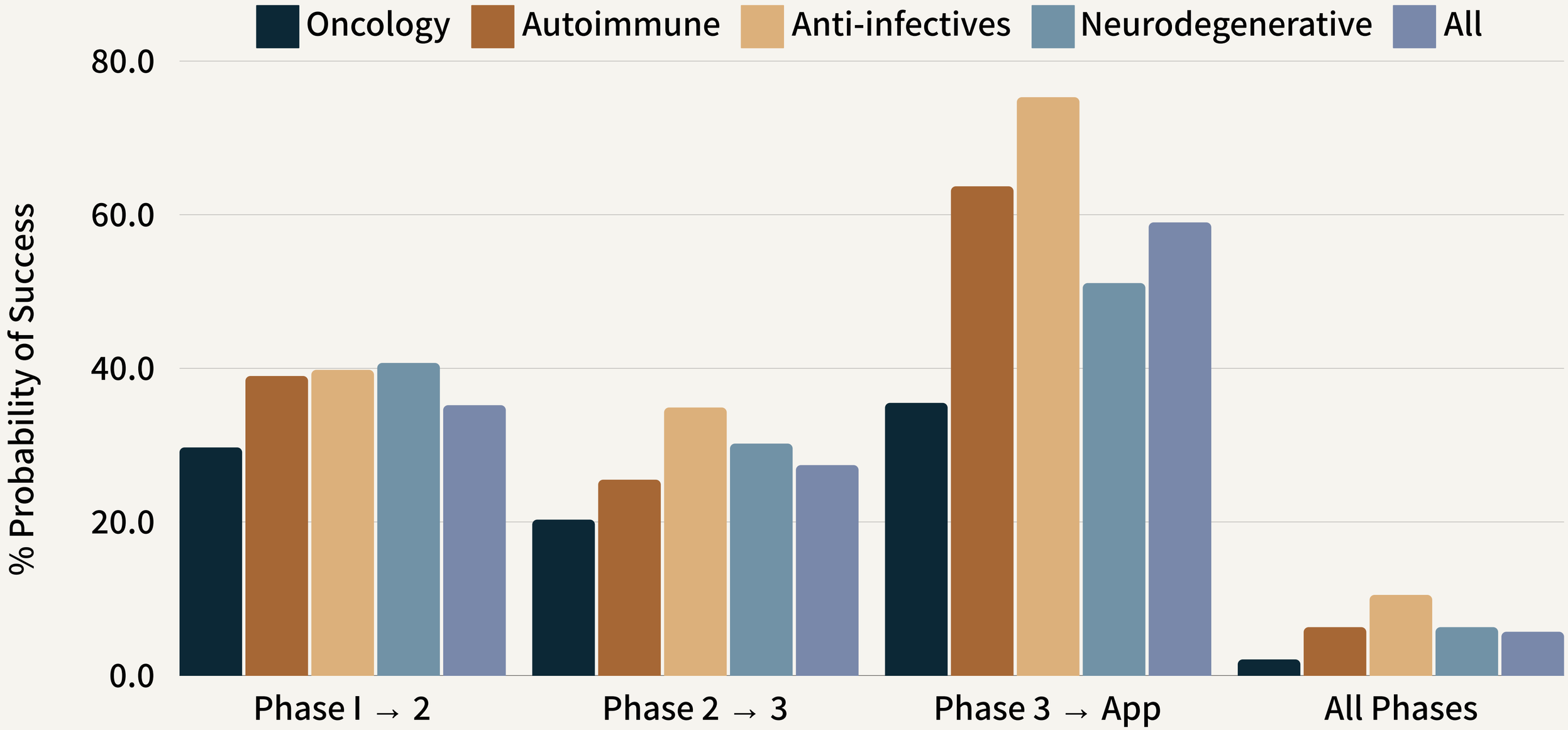
Emulate Bio

Predictive Toxicology

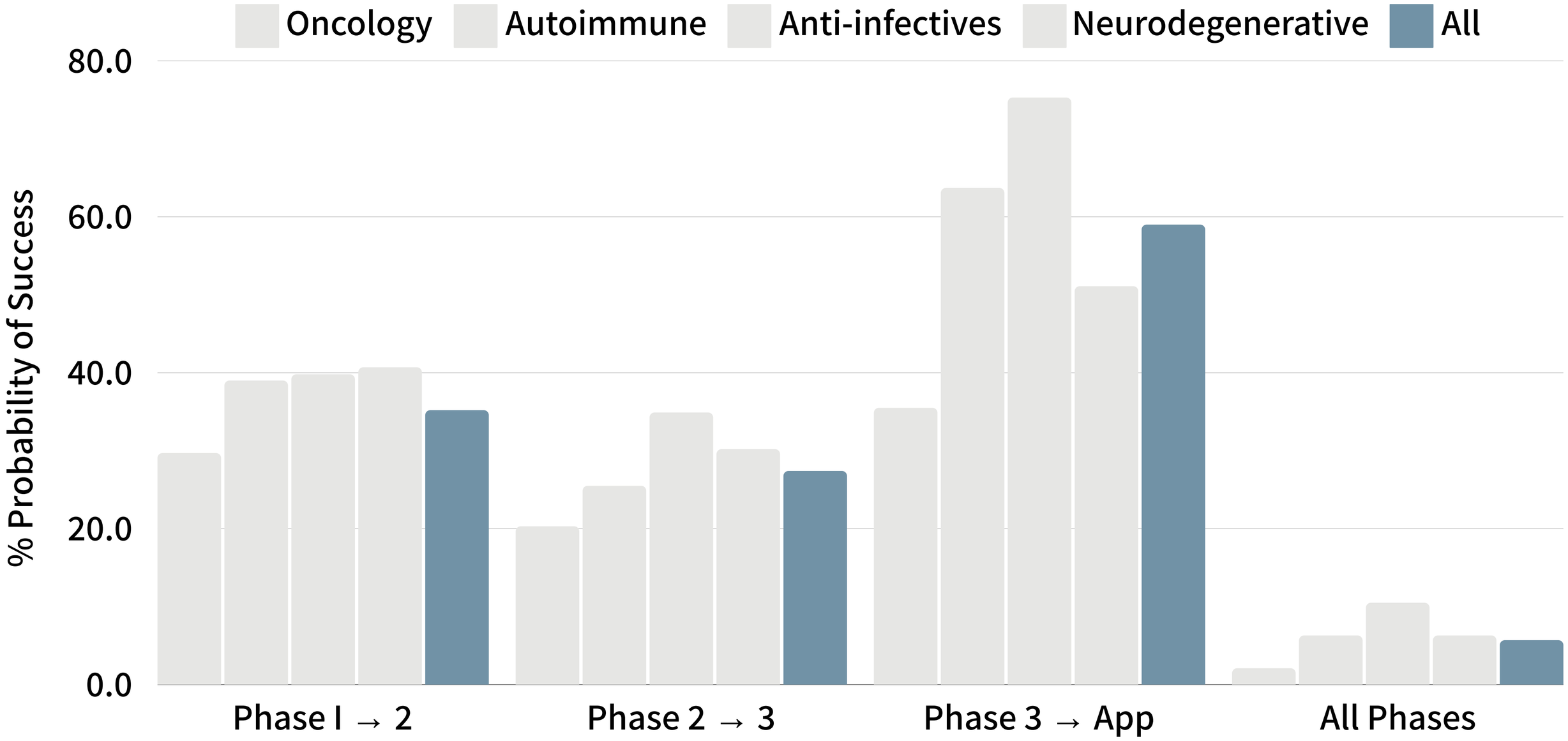
- **Tox21**
- **eTOX**
- **OASIS**
- **IQ-DILI**

Wellcome Leap: HOPE

Probability of Success | Disease areas

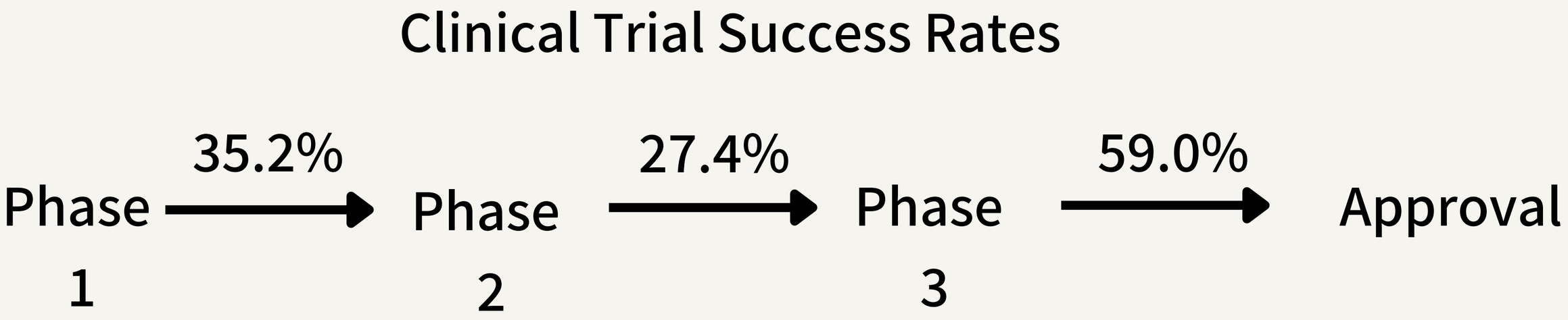


Source data: Adapted from Wong, et. al. (2019)., Biostatistics



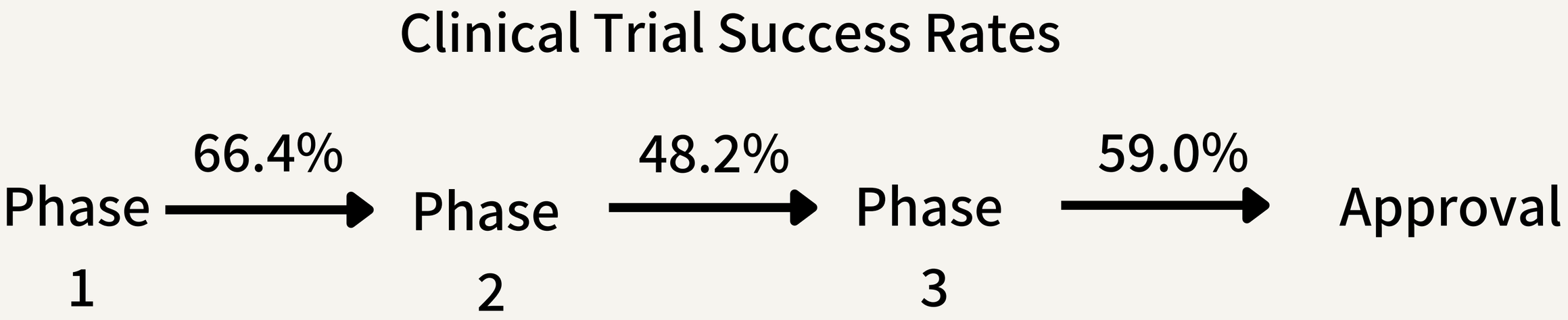
Source data: Adapted from Wong, et. al. (2019)., Biostatistics

Drug development estimated success rate of 5.7%



Source data: Adapted from Wong, et. al. (2019)., Biostatistics

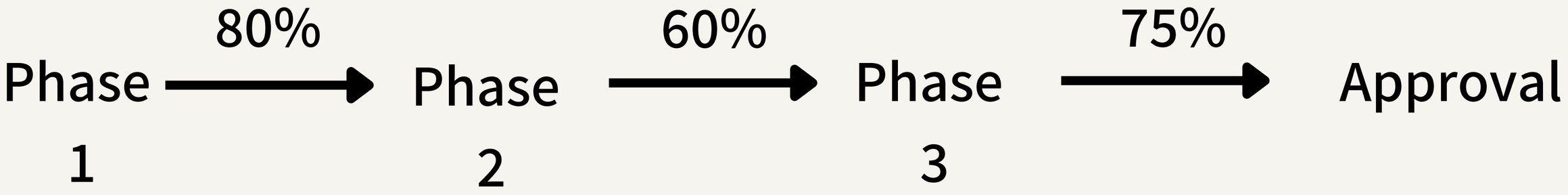
Drug development estimated success rate of 13% (Alt method)



Source data: Adapted from Wong, et. al. (2019)., Biostatistics

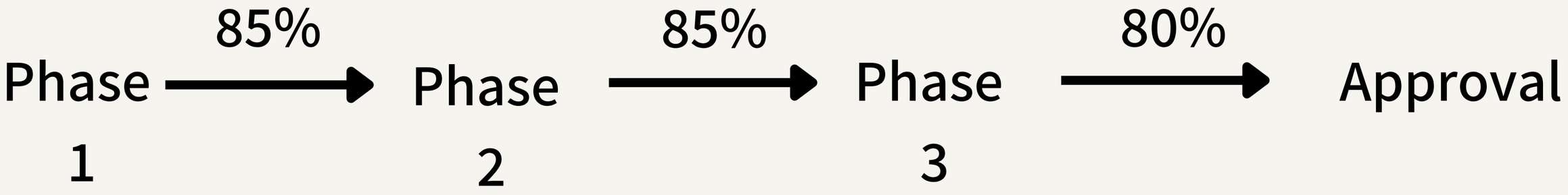
5x success rates in < 5 years

36% overall success rates in a single disease area



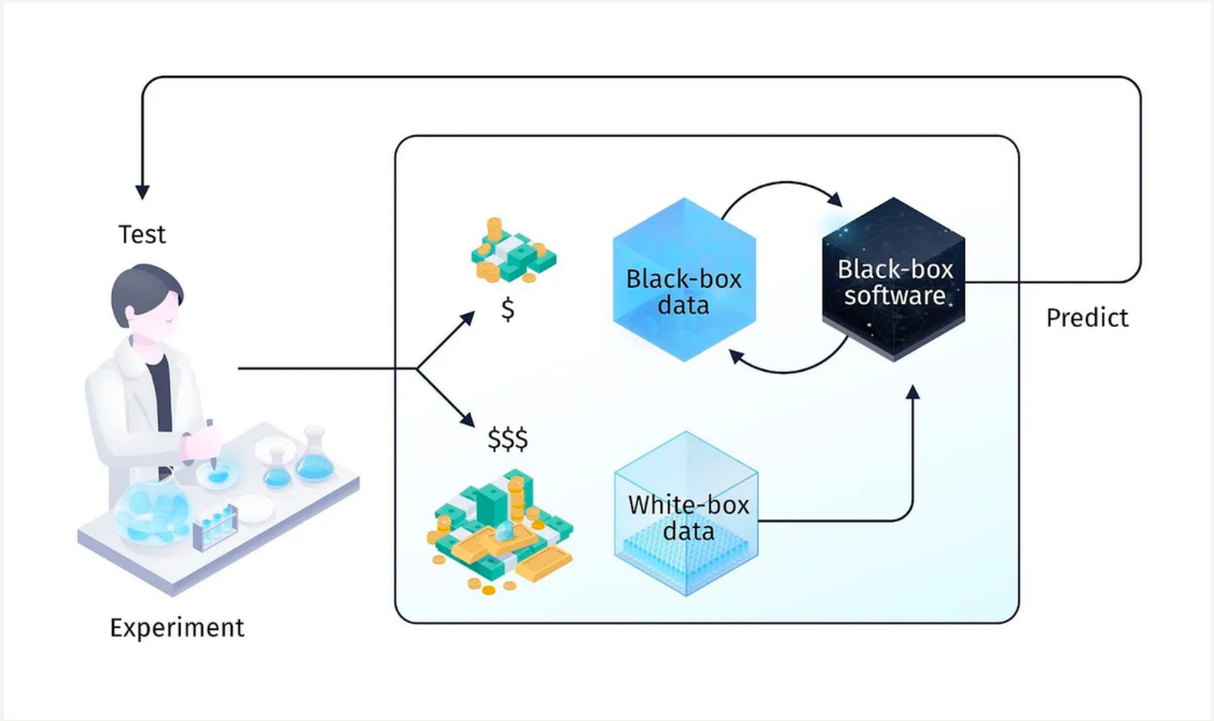
10x success rates in 10 years

57% overall success rates across all diseases



ATHENA vs. Other AI: Importance of High Quality Data

“Generation 3” AI-driven biological science.



Credit: Bronstein and Naef. "Road to Biology 2.0"

“

… requirement to have **sufficient paired “white-box” + “black-box” data to train and evaluate** such a model.

Michael Bronstein,
DeepMind Professor of AI, University of Oxford.

“ I think the best way to make progress on applications of machine learning to drug discovery is to **fund a large public effort that will generate high-quality data and make this data available to the community.**

…

It frustrates me that funding agencies continue to support research into new ML methods but **are unwilling to fund efforts to generate data that can be used to train and validate ML models for drug discovery.**

Patrick Walters,
Chief Data Officer of Relay Therapeutics

https://practicalcheminformatics.blogspot.com/2023/08/we-need-better-benchmarks-for-machine.html

“

A key point that cannot be overemphasized: **We lack consistent and quality data for both training and validation across most applications in drug discovery.**

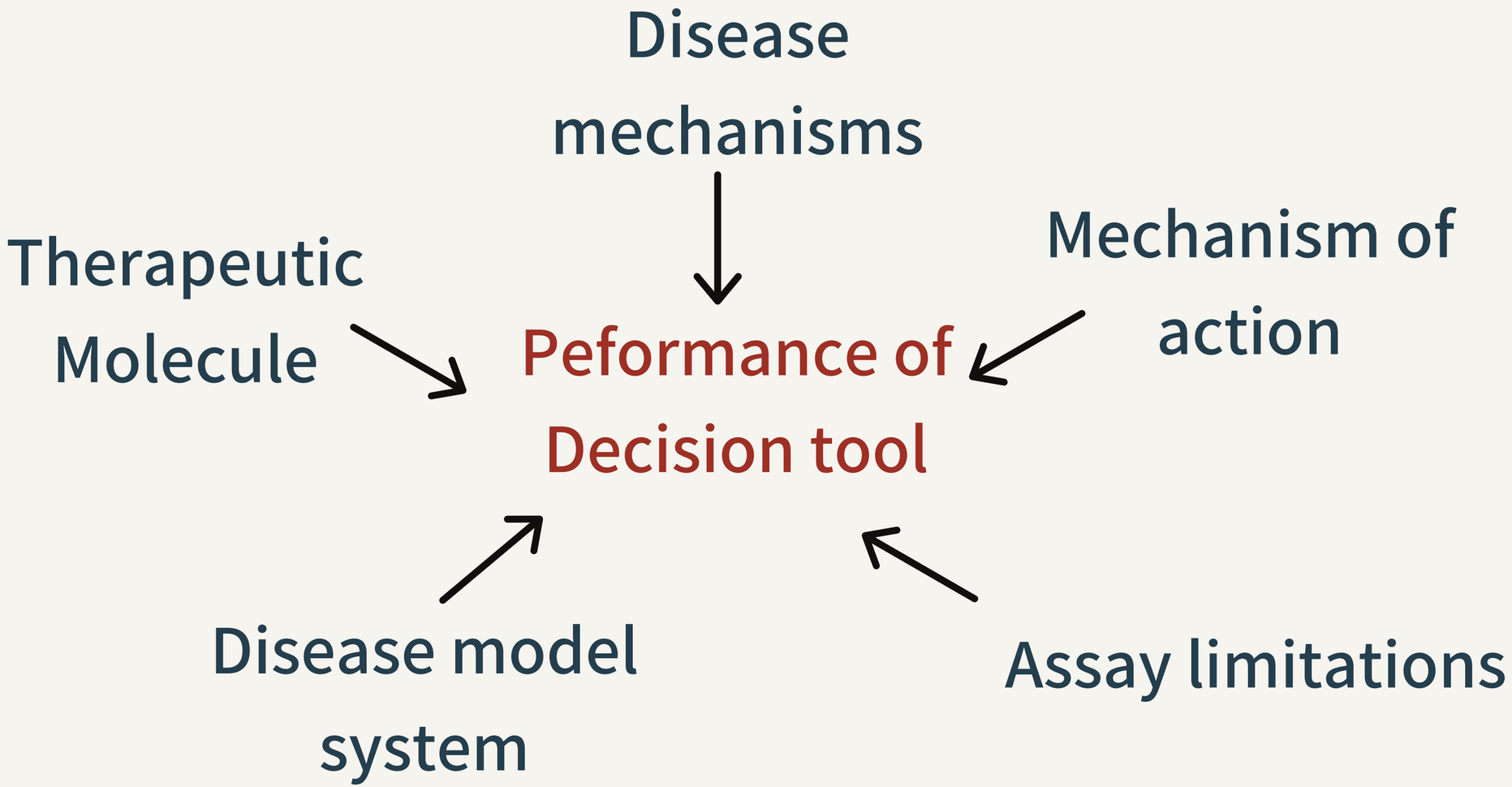
Sam Bjork & Travis Hughes
Digitalis Ventures

“

Generally too few data to expect algorithms to infer relationships between drug treatment and in vivo endpoint observations

Andreas Bender,
Professor of Molecular Informatics, Cambridge

What is Causal AI?


Disease mechanisms
Therapeutic Molecule
Mechanism of action
Peformance of Decision tool
Disease model system
Assay limitations

Why causal inference?

1

Invariant causal representations

2

Account for measurement ≠ biological construct

3

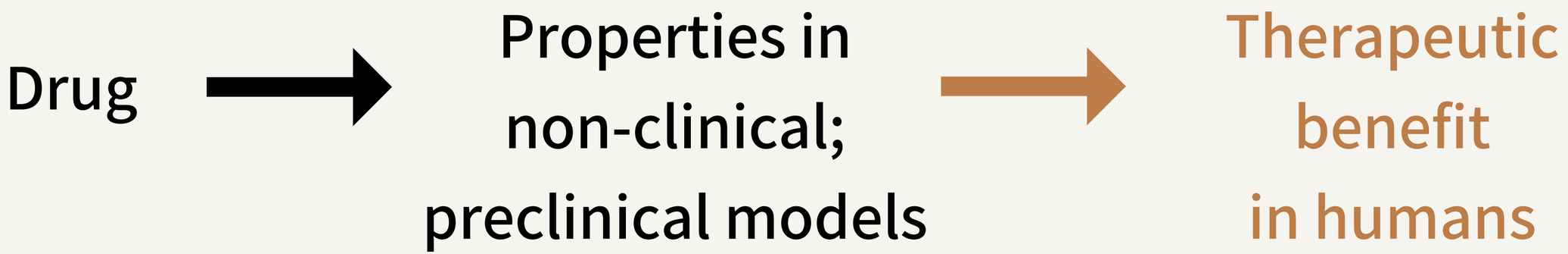
Evidential Triangulation: fuse orthogonal evidence across studies

Labels are highly ambiguous — biology is context dependent

Drug  Properties in non-clinical; preclinical models

f(X, y)

Labels are highly ambiguous — biology is context dependent



f(X, y)

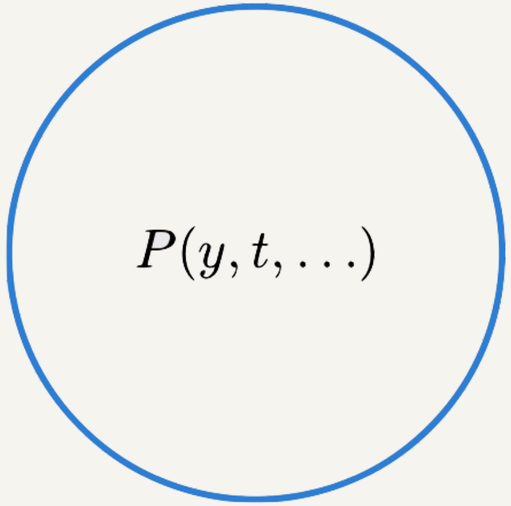
Training

f(X*, y*)

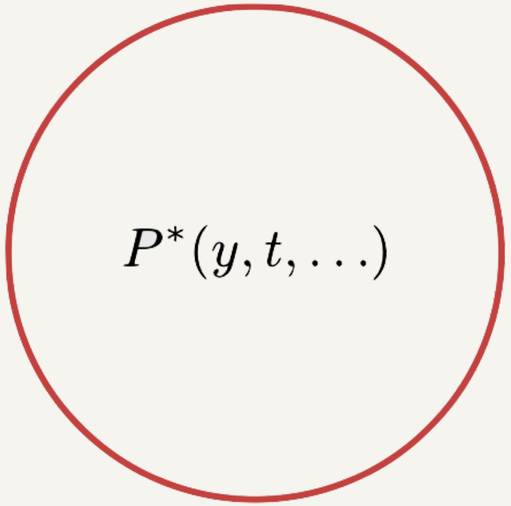
Testing

Transportability problem

Source Population Π



Target Population Π^*



Given $P(y \mid do(t), x)$

$P(y \mid do(t), x) \stackrel{?}{=} P^*(y \mid do(t), x)$

Bareinboim, Elias, and Judea Pearl. 2016. “Causal Inference and the Data-Fusion Problem.” Proceedings of the National Academy of Sciences of the United States of America 113 (27): 7345–52.

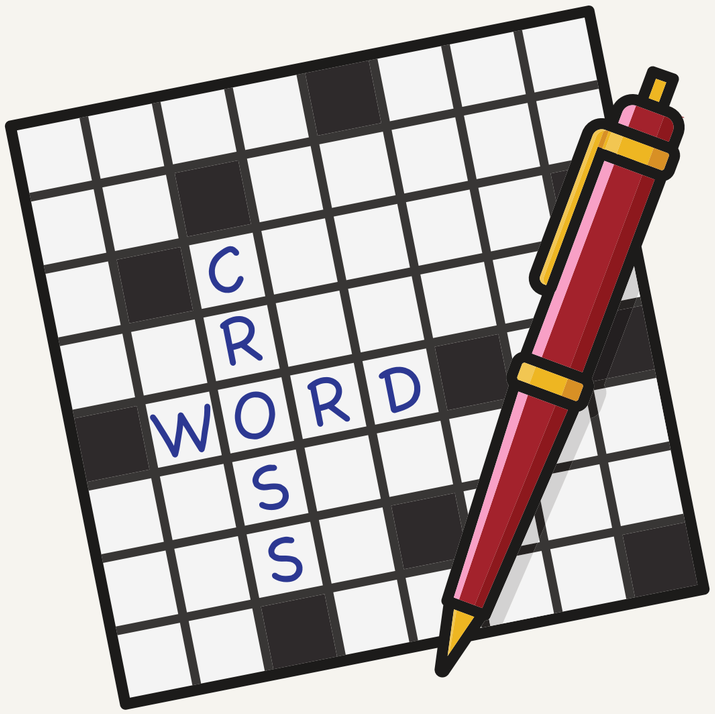
Lee, Sanghack, Juan Correa, and Elias Bareinboim. 2020. “General Transportability – Synthesizing Observations and Experiments from Heterogeneous Domains.” Proceedings of the AAAI Conference on Artificial Intelligence 34 (06): 10210–17.

Breskin, Alexander, Stephen R. Cole, Jessie K. Edwards, Ron Brookmeyer, Joseph J. Eron, and Adimora A. Adimora. 2021. “Fusion Designs and Estimators for Treatment Effects.” Statistics in Medicine 40 (13): 3124–37.

Dahabreh, Issa J., Anthony Matthews, Jon A. Steingrimsson, Daniel O. Scharfstein, and Elizabeth A. Stuart. 2023. “Using Trial and Observational Data to Assess Effectiveness: Trial Emulation, Transportability, Benchmarking, and Joint Analysis.” Epidemiologic Reviews, February. https://doi.org/10.1093/epirev/mxac011.

Squires, Chandler, Dennis Shen, Anish Agarwal, Devavrat Shah, and Caroline Uhler. 11--13 Apr 2022. “Causal Imputation via Synthetic Interventions.”. Proceedings of Machine Learning Research. PMLR.

Evidential Triangulation

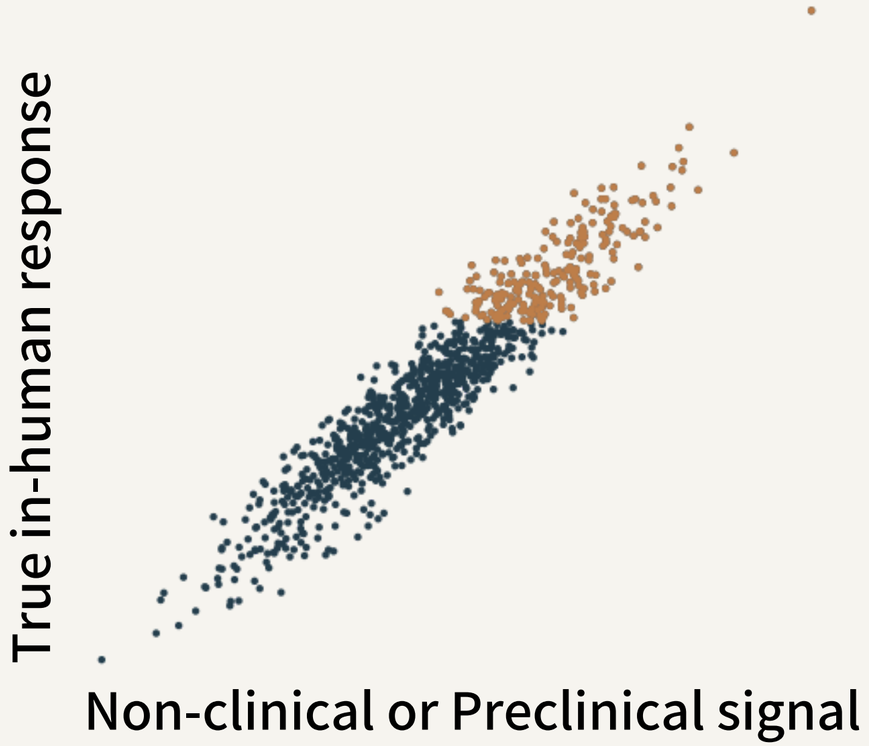


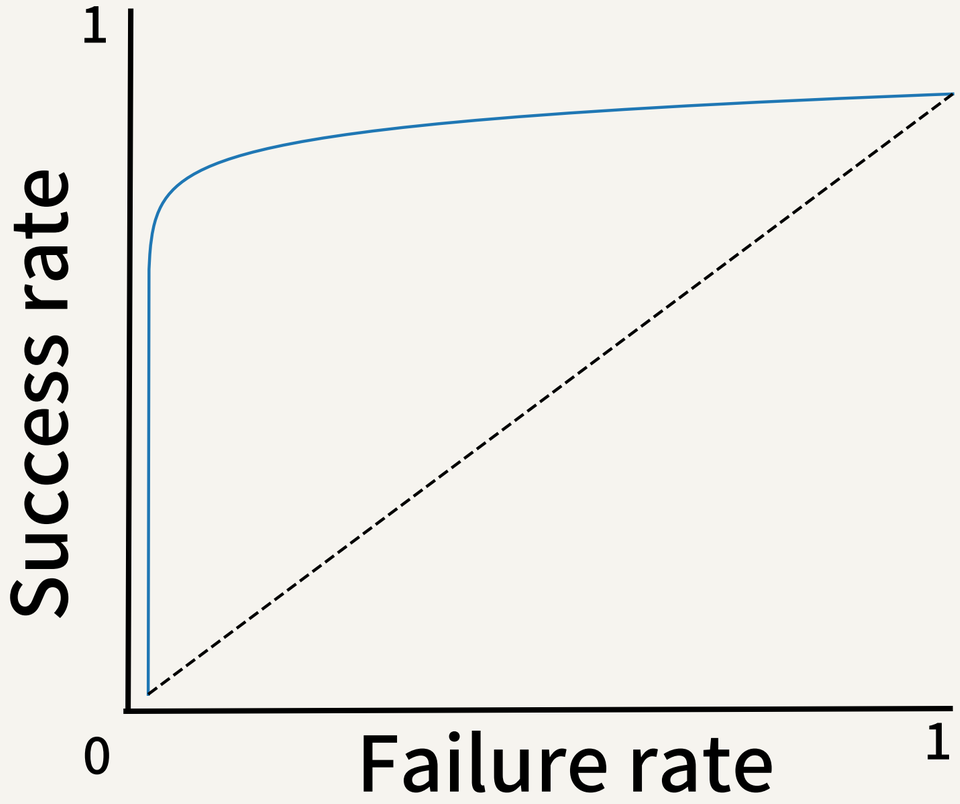
- Clues — partially accurate tool
- Words — in-human predictions per tool

- Borrowing strength — Identify tools with orthogonal biases
- Triangulation — interlocking of independently confident words

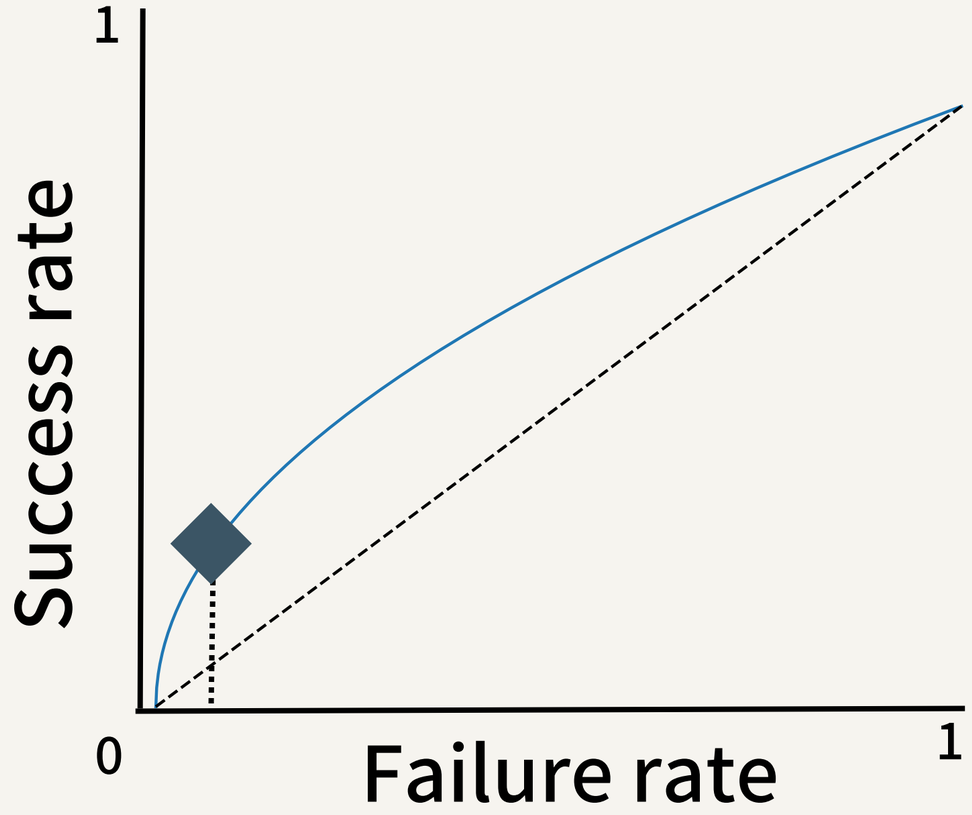
Decision Tools

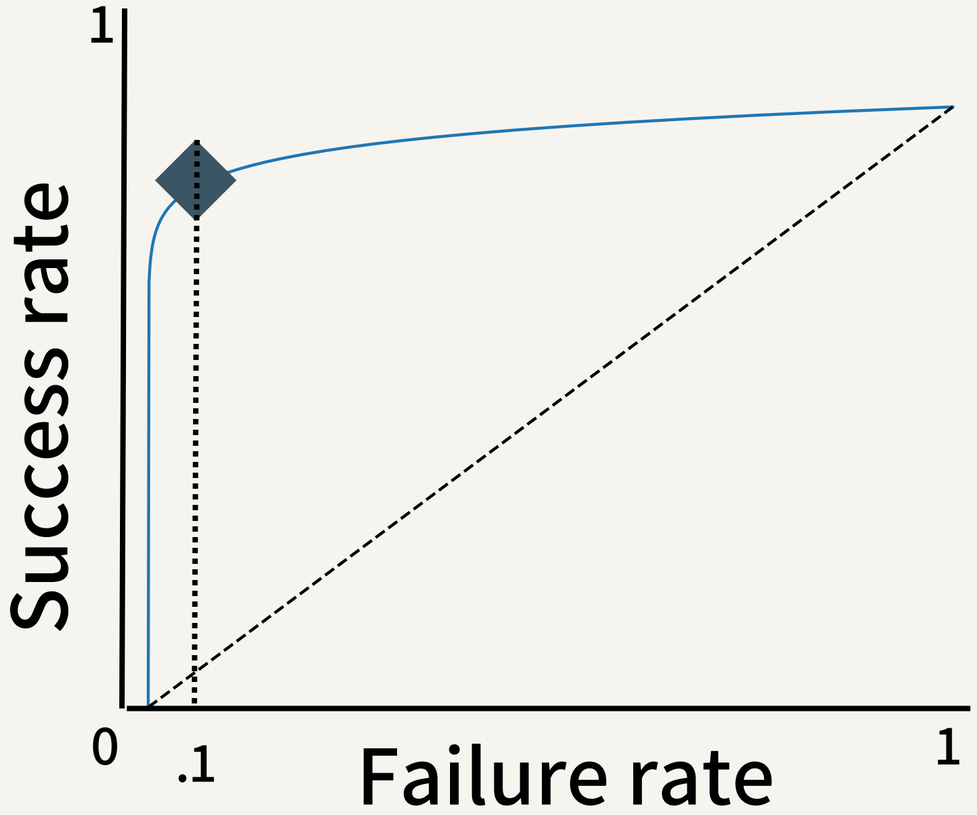
Operating Characteristics of Proxy Signals



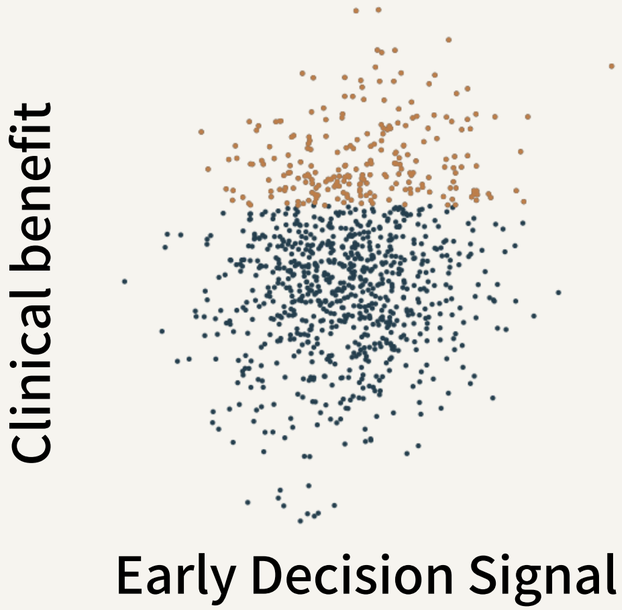


Not about reducing false positives





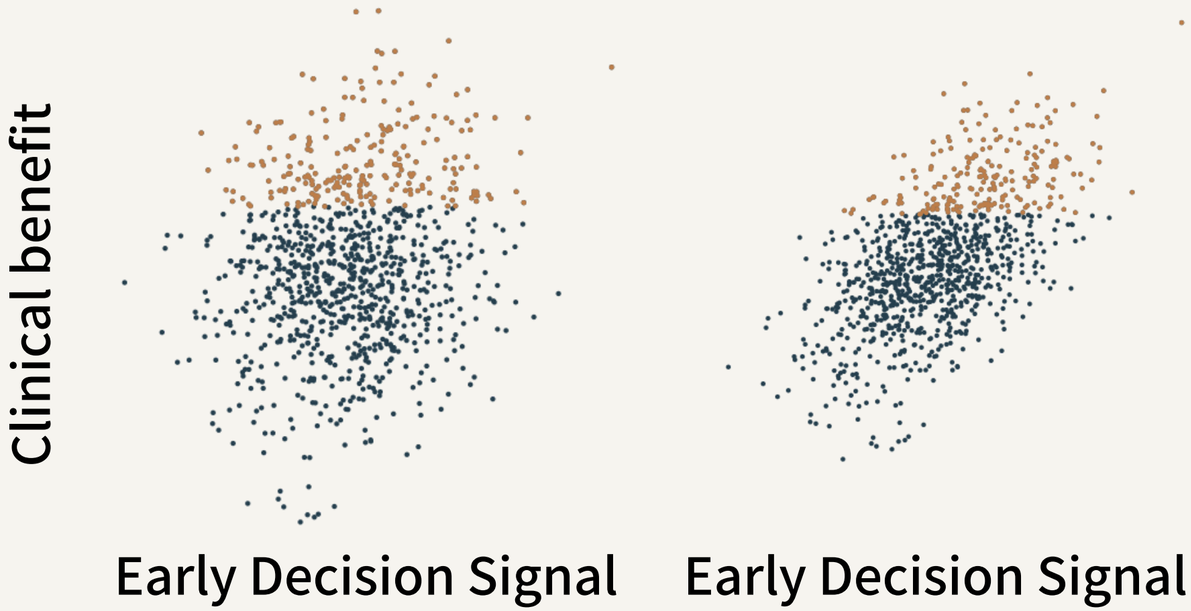
Quality of decision tools matter!



Validity implies ability to distinguish winners from losers

Source data: Adapted from Scannell. et. al. (2022)., Nat. Rev. Drug. Discov.

Quality of decision tools matter!



Difference between low validity (.2) to moderate validity (.5) for phase 1 is worth >100m

Source data: Adapted from Scannell. et. al. (2022)., Nat. Rev. Drug. Discov.

Examples from Oncology Decision Tools

The case of cancer drug development

Success rate of **~2.1%**

Costs of 4 billion per novel drug entity

Source data: Adapted from Wong, et. al. (2019)., Biostatistics

Low validity of decision tools leads to poor translational success – 1

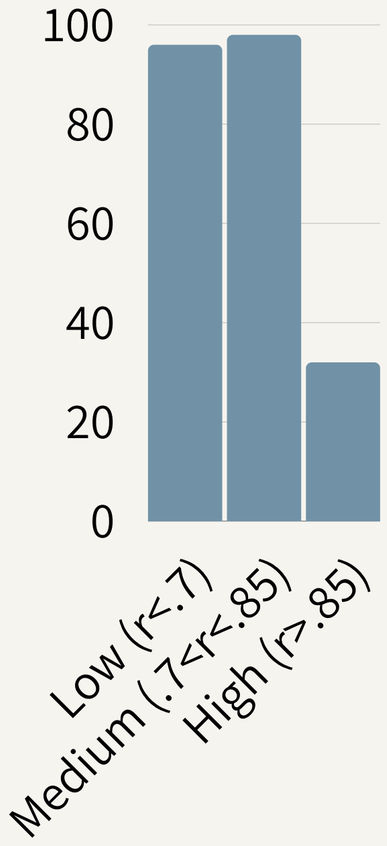
Weak precision biomarkers

→

fail to separate who benefits from immune checkpoint inhibitors from antibody drug conjugates or chemotherapy

Source data: Adapted from Prasad, et. al. (2015)., JAMA Intern Med. doi:10.1001/jamainternmed.2015.2829

Low validity of decision tools leads to poor translational success – 1



Poorly validated biomarkers for surrogate endpoints

→

treatments do not improve overall survival (paclitaxel+bevacizumab)

Source data: Adapted from Prasad, et. al. (2015)., JAMA Intern Med. doi:10.1001/jamainternmed.2015.2829

Low validity of decision tools leads to poor translational success – 1

"The use of surrogate end points in oncology has led to use of toxic drugs that do not improve survival. For example, **bevacizumab (Avastin; Roche/Genentech) gained FDA accelerated approval for metastatic breast cancer in 2008 based on evidence that it could improve PFS.** In 2011, that approval was withdrawn when multiple randomized trials confirmed that bevacizumab did not improve overall survival and that gains in PFS were more modest than those seen in the earlier trial. Nevertheless, Medicare and other insurers are still obliged to pay for bevacizumab because it remains endorsed by the NCCN for this indication."

Source data: Adapted from Prasad, et. al. (2015)., JAMA Intern Med. doi:10.1001/jamainternmed.2015.2829

Low validity of decision tools leads to poor translational success – 2

In-vitro & in-vivo drug response prediction of anti-cancer drugs show poor corroboration to phase 1 trials for safety

Source data: Bae et al. Cell Death and Disease (2020). doi.org/10.1038/s41419-020-2462-8;
Jenzch et. al. JAMA Netw Open.(2023) doi:10.1001/jamanetworkopen.2023.24977

Low validity of decision tools leads to poor translational success – 2

Preclinical disease models of therapeutic benefit don’t translate to humans

→

Failure of IGF-1R inhibitors in clinical settings

Source data: Bae et al. Cell Death and Disease (2020). doi.org/10.1038/s41419-020-2462-8;
Jenzch et. al. JAMA Netw Open.(2023) doi:10.1001/jamanetworkopen.2023.24977

Low validity of decision tools leads to poor translational success – 2

“Xenograft data considered to be predictive of the successful clinical activity of IGF-1R inhibitors, all of which subsequently failed to show clinical efficacy, can either be considered to invalidate the models themselves (when there was 100% tumor growth inhibition but no clinical activity) or, at the least, the interpretation of them. In 2014, drug researchers at AstraZeneca articulated a 5-dimensional framework that reduced their drug attrition in clinical trials: more rigorous target validation and improved preclinical models were key aspects that required attention.”

Jenzch et. al. JAMA Netw Open.(2023) doi:10.1001/jamanetworkopen.2023.24977