Drugs don't care about ‘AI’ - But AI needs to care about drugs

Andreas Bender, PhD
Professor for Machine Learning in Medicine, Khalifa University, Abu Dhabi, UAE

Visiting Professor, Yusuf Hamied Department of Chemistry, University of Cambridge, UK
Research Professor at UBB and Project Leader at UMF, Cluj-Napoca, Romania
Adjunct Faculty at National Institute for Bioprocessing Research and Training (NIBRT), Dublin, Ireland
Co-Founder of Healx, Ltd., PharmEnable, Ltd., and Pangea Bio Ltd.

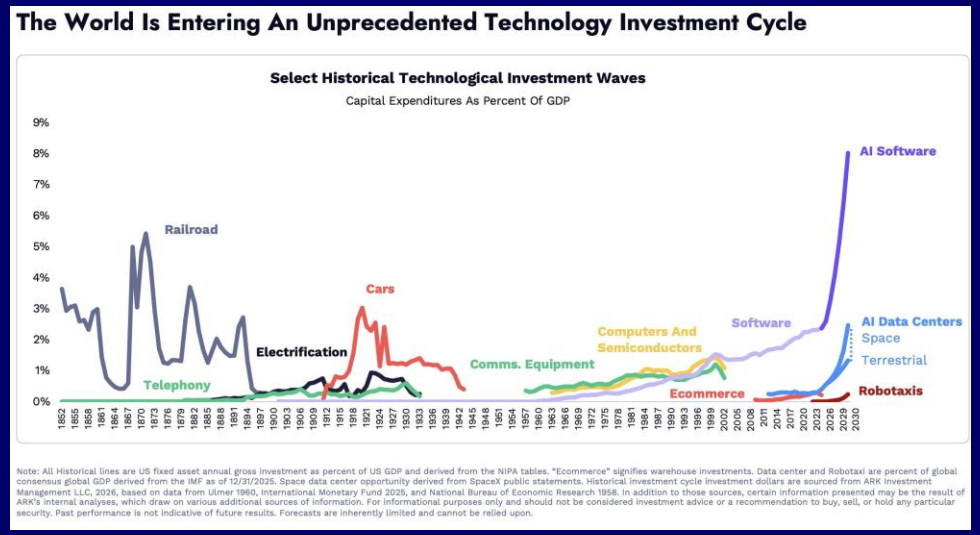
جامعة خليفة
Khalifa University

Slides at: www.drugdiscovery.net/broad26

Currently we see a broad ‘AI Disconnect’ – e.g.

Currently we see a broad ‘AI Disconnect’ – e.g.

Investment in AI as fraction of GDP beyond that on railroads *etc.*

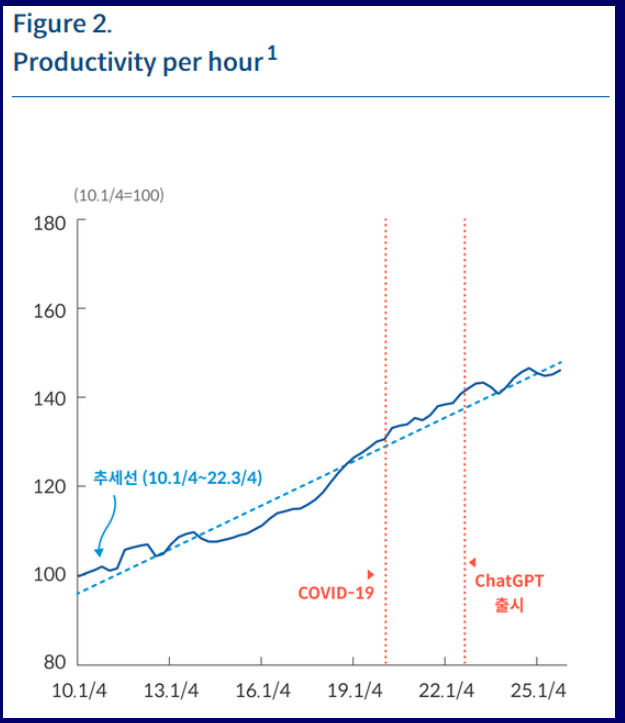


VS

Marginal impact on productivity, e.g. Bank of Korea estimates ~1% impact on productivity due to AI (similar recent report by BoE)

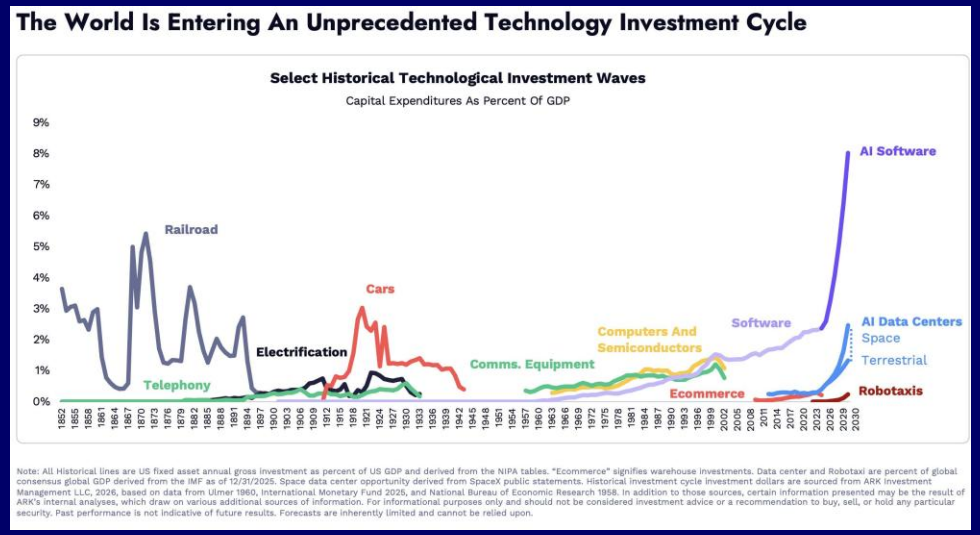
Does AI increase productivity?

Bureau of Employment Research Team
Chief Seo Dong-hyun, Team Leader Oh Sam-il, Research Station Yoon Jong-won
2026.06.18



Currently we see a broad ‘AI Disconnect’ – e.g.

Investment in AI as fraction of GDP beyond that on railroads *etc.*



VS

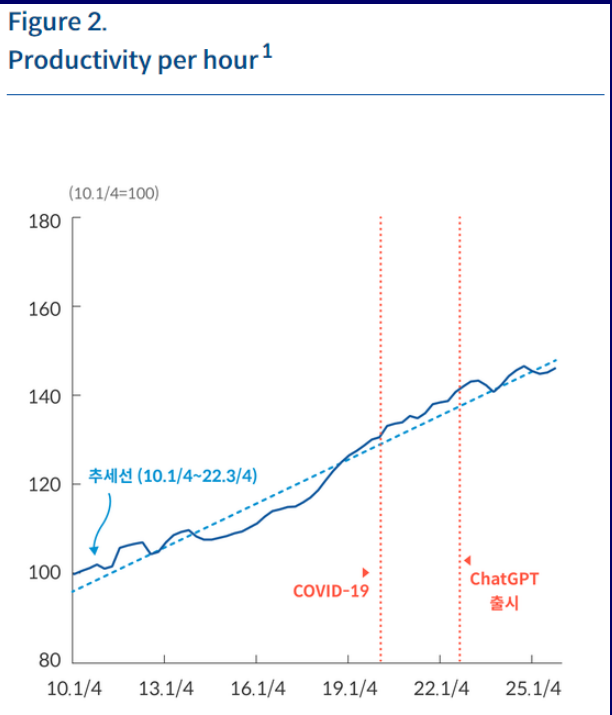
Marginal impact on productivity, e.g. Bank of Korea estimates ~1% impact on productivity due to AI (similar recent report by BoE)

Does AI increase productivity?

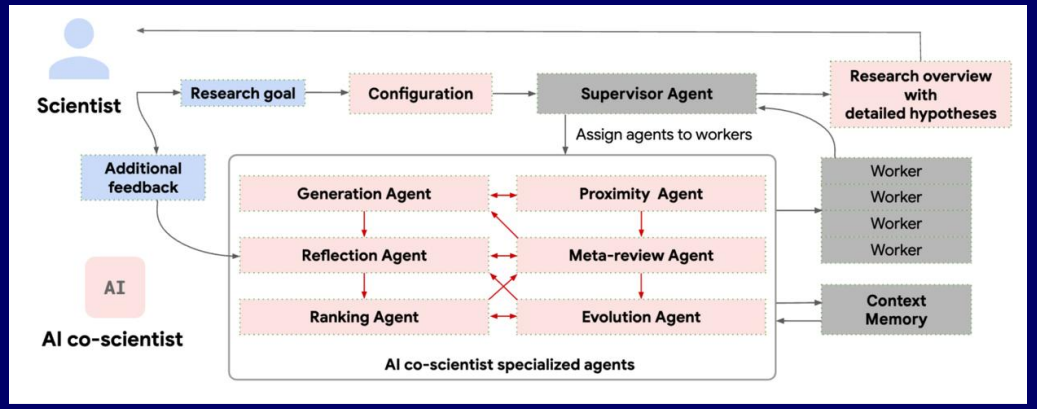
Bureau of Employment Research Team

Chief Seo Dong-hyun, Team Leader Oh Sam-il, Research Station Yoon Jong-won

2026.06.18

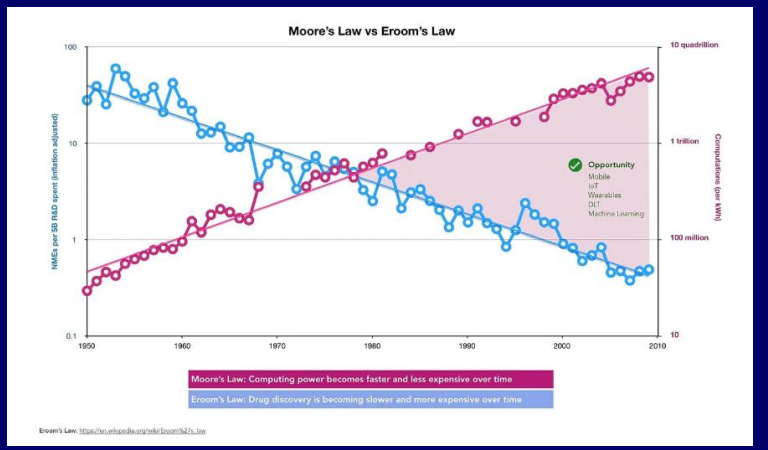


End-to-end learning, agents, foundation models, virtual cells, …

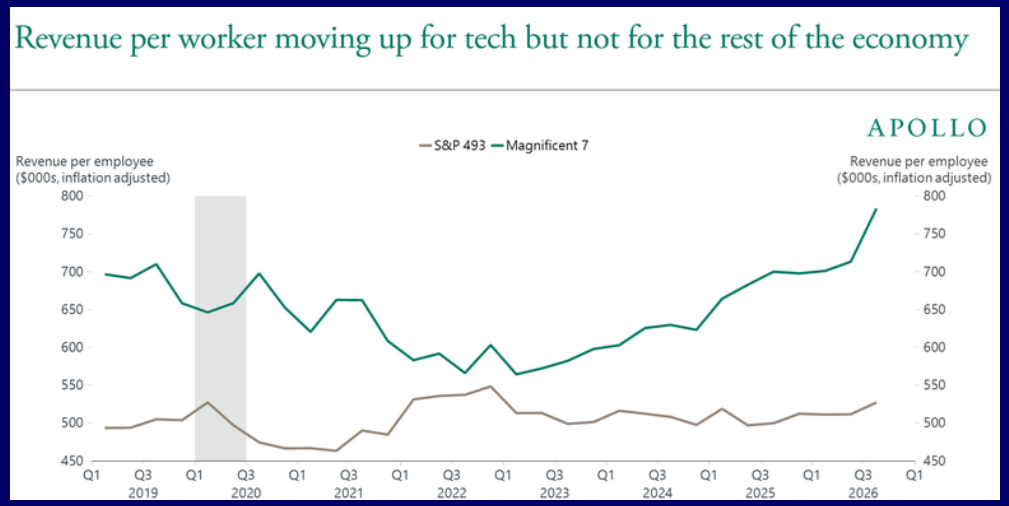


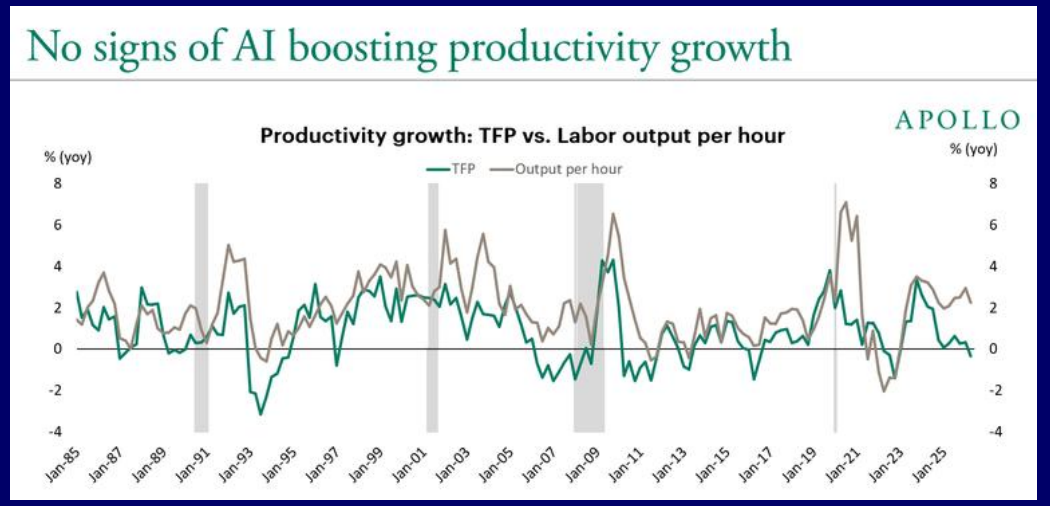
VS

Eroom’s Law in drug discovery (inverse of Moore’s Law)



AI does not (*currently*) increase revenue, productivity, profit margins, … *outside tech*, that is

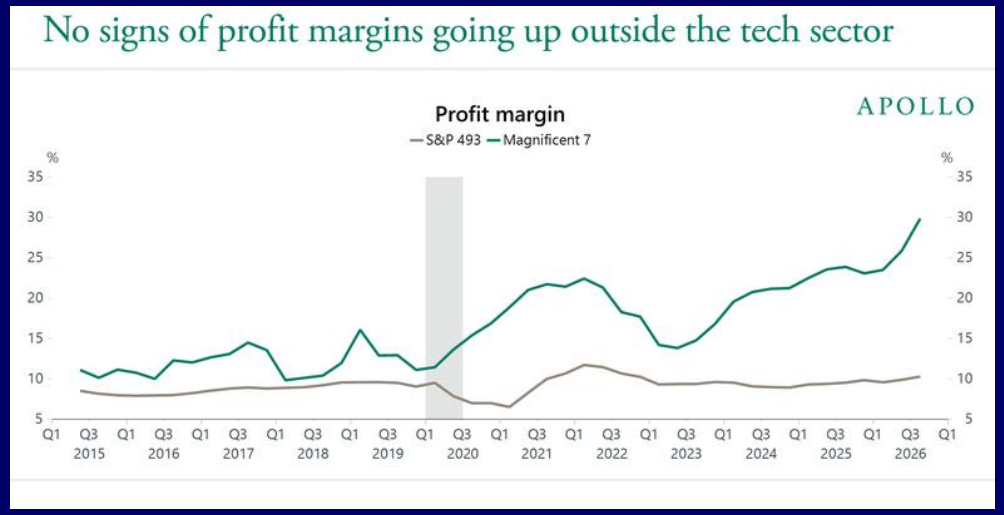


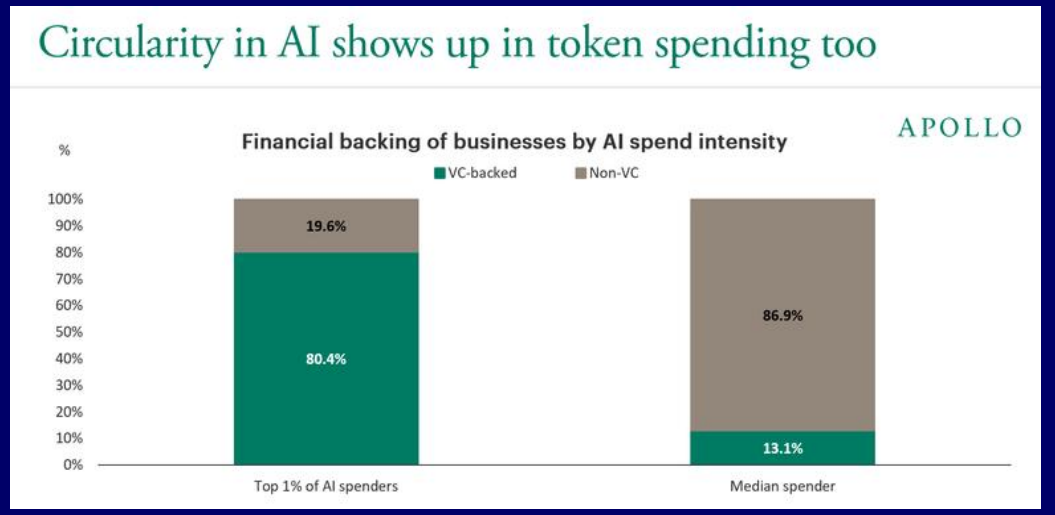


MACROECONOMIC INDICATORS & TRENDS

October 24, 2026 APOLLO

No Signs of AI Payoff Outside the Tech Sector





https://www.apollo.com/wealth/insights-news/insights/daily-spark/no-signs-of-ai-payoff-outside-the-tech-sector

Personal observation: Decrease of signal-to-noise ratio (*precision)* everywhere

Every start-up tells you that…

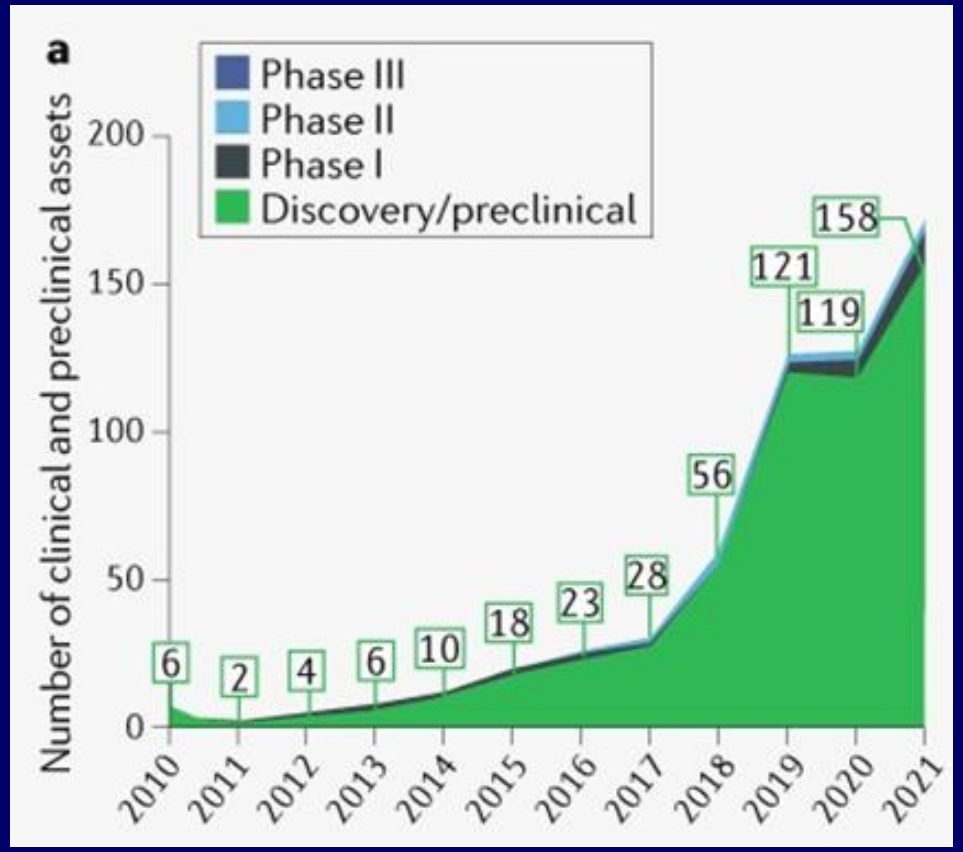
'Our model achieves 93% Performance on This and that Benchmark, which Outperforms SOTA and revolutionizes drug Discovery, for the 1001st time'

SOTA = ‘State Of The Art’, a term frequently used in machine learning that something is as good as it currently gets

… but we see little translation into the clinic, and clinical success

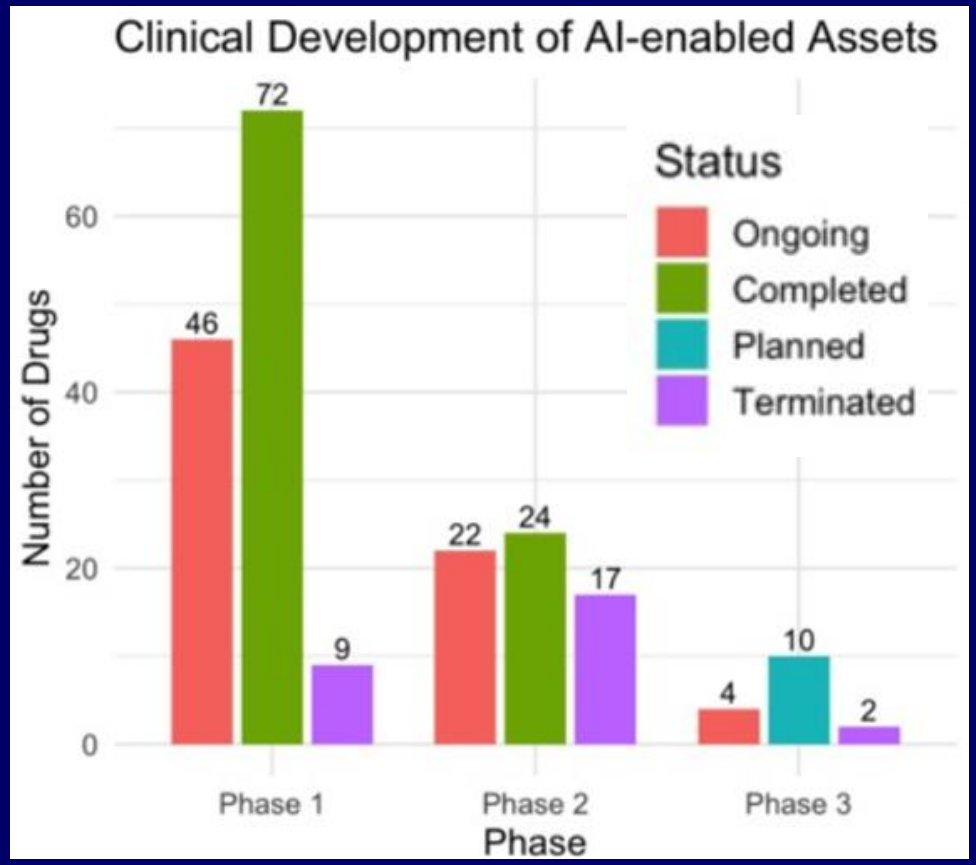
Significant *number* of *discovery* programs of ‘AI-native’ companies

Very little Ph1, less Ph2, 1 in Ph3 (2023 data)



Jayatunga et al., AI in small-molecule drug discovery: a coming wave? *Nature Rev. Drug Disc.* 2022; June 2023 Wellcome/BCG Report “Unlocking the Potential of AI in Drug Discovery”

Note: Virtually every current therapy had an ‘AI’ component during its genesis, and it is very difficult to obtain clean related data



3/2026 data (129 assets, 69 companies): Still few successful late-stage clinical successes reported

Haque et al., Artificial intelligence–enabled drug discovery: A growing but largely early-stage clinical pipeline *Cancer Treat. Rev.* 2026 (151) 103223.

The macro view: Available data – *up*; FDA approvals – flat … so does all that data really help us?!

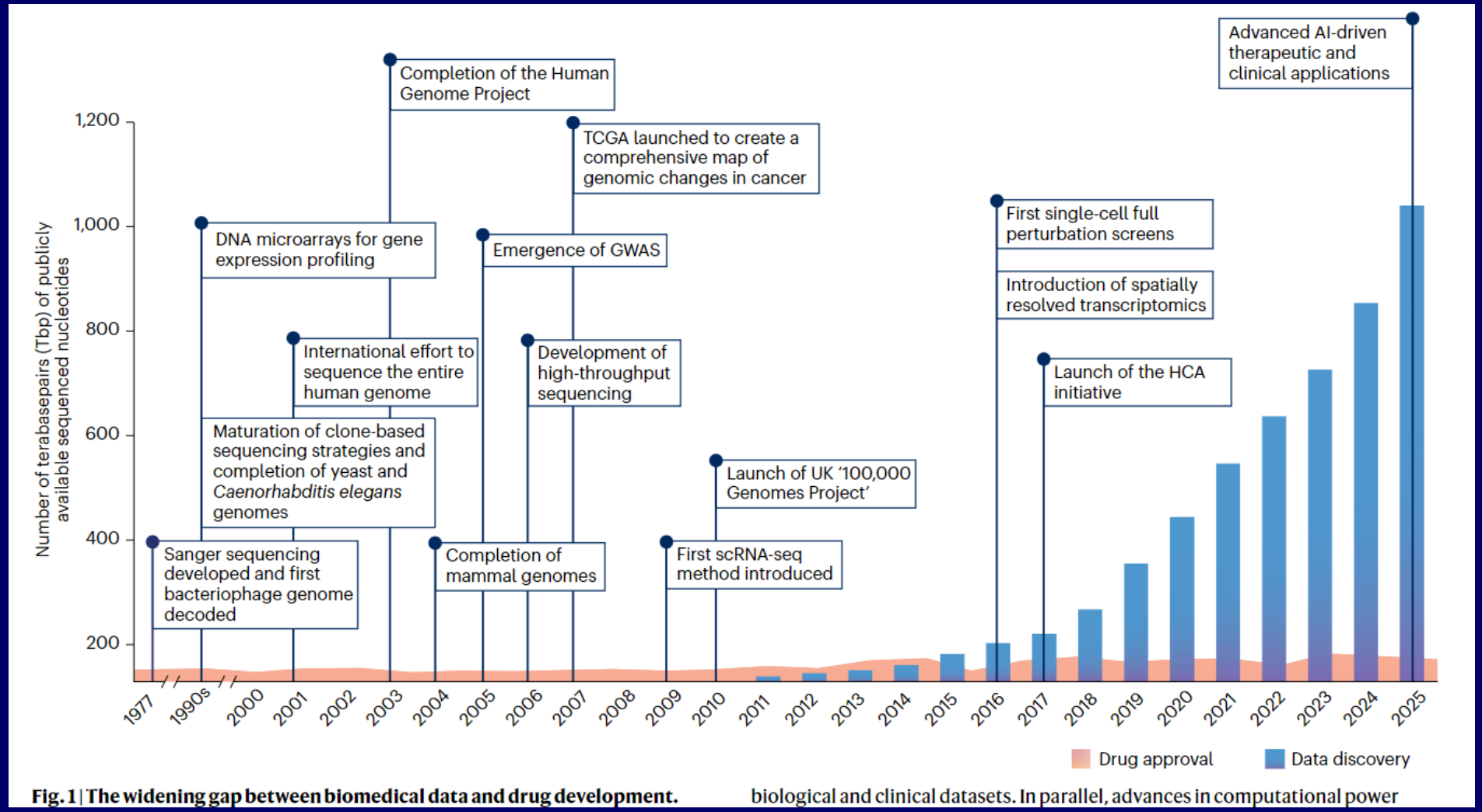
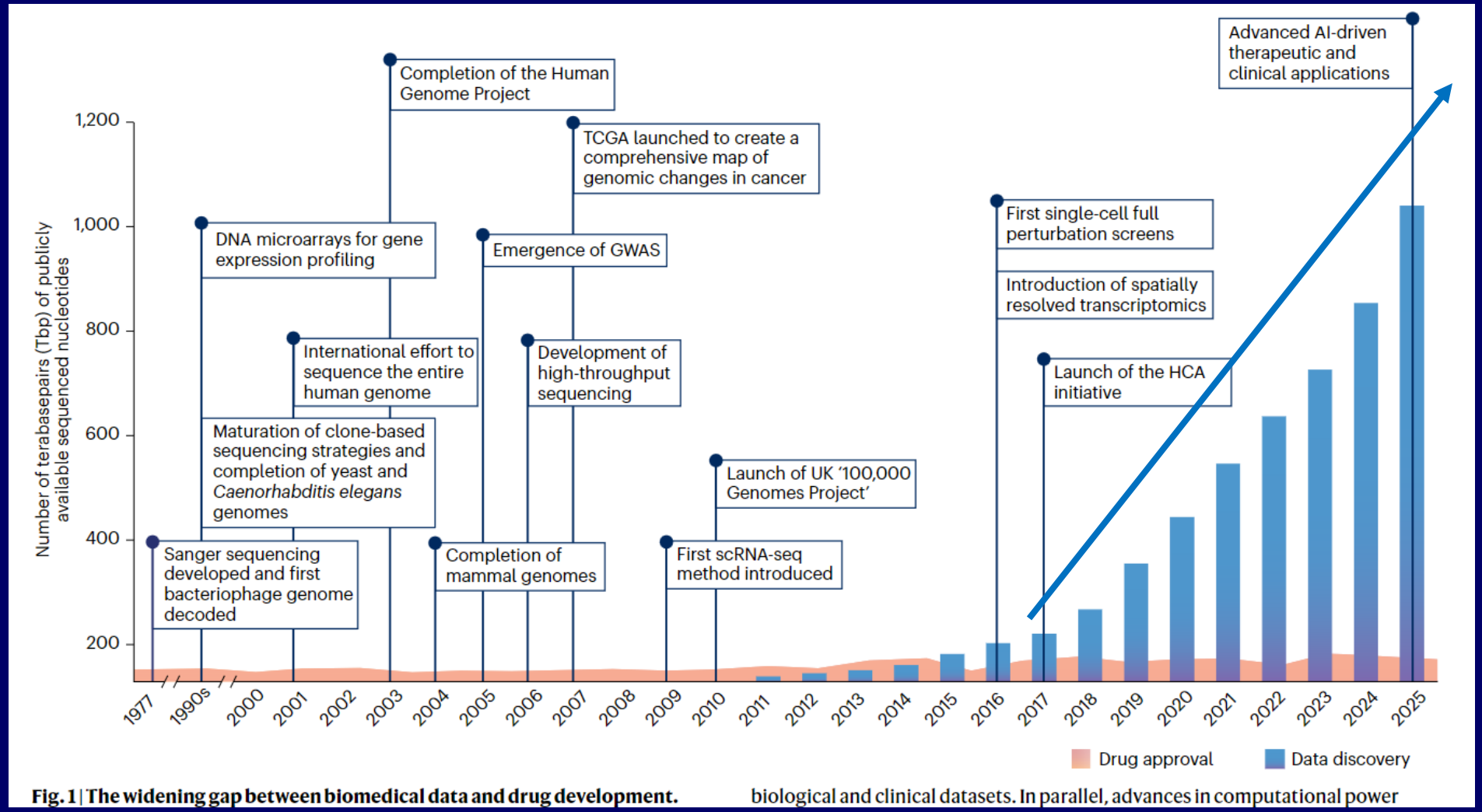


Fig. 1 | The widening gap between biomedical data and drug development. biological and clinical datasets. In parallel, advances in computational power

Shalita & Amit, Nature Biomed Eng. 2026 https://doi.org/10.1038/s41551-026-01785-6

The macro view: Available data – *up*; FDA approvals – flat … so does all that data really help us?!



DATA

Fig. 1 | The widening gap between biomedical data and drug development. biological and clinical datasets. In parallel, advances in computational power

Shalita & Amit, Nature Biomed Eng. 2026 https://doi.org/10.1038/s41551-026-01785-6

The macro view: Available data – *up*; FDA approvals – flat … so does all that data really help us?!

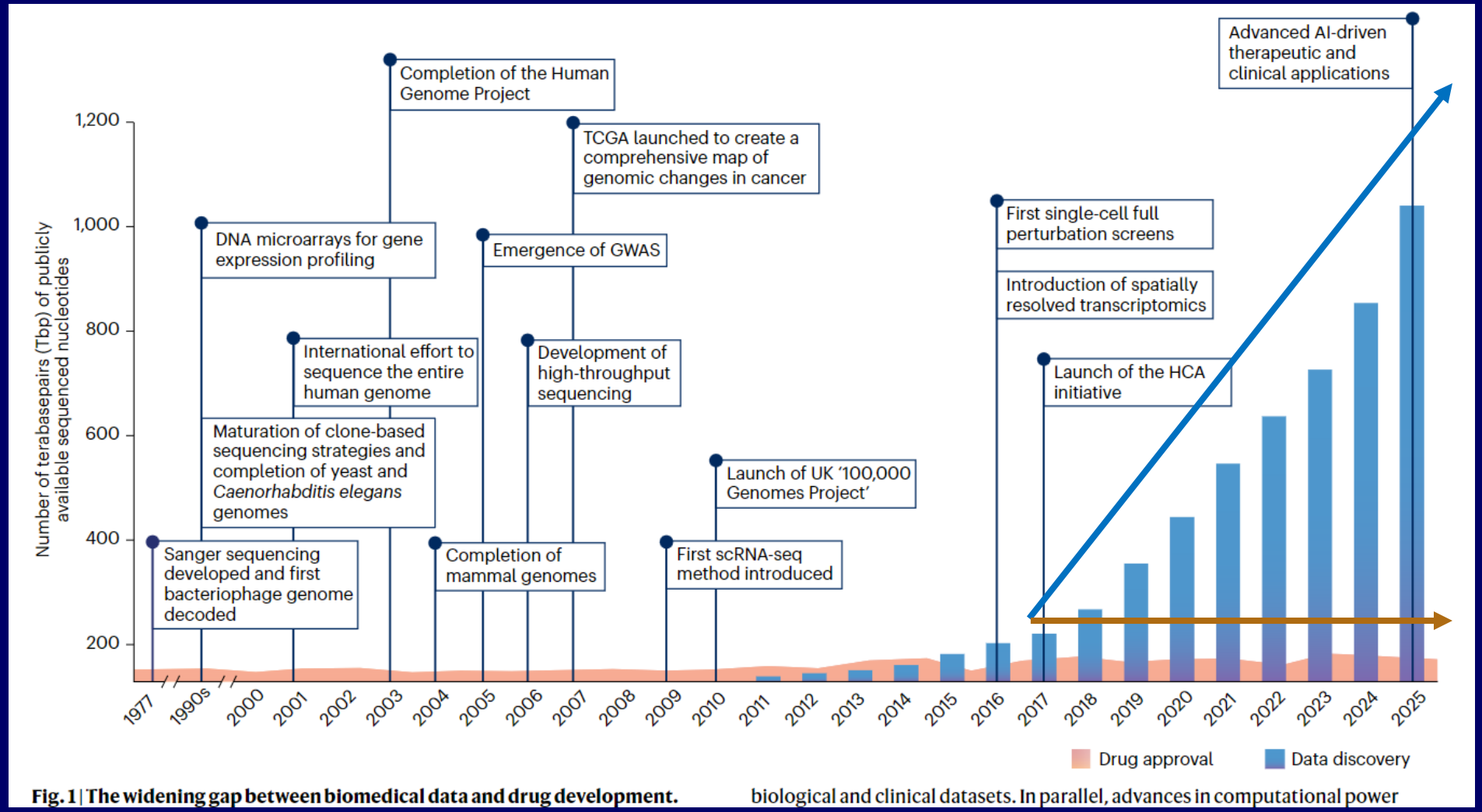


Fig. 1 | The widening gap between biomedical data and drug development. biological and clinical datasets. In parallel, advances in computational power

DATA

APPROVALS

Shalita & Amit, Nature Biomed Eng. 2026 https://doi.org/10.1038/s41551-026-01785-6

Let’s think a bit – about our goal, models, ‘validation’, …

1. What does it mean to ‘discover’ a drug?
2. Aspects of model validation
3. On data and decision spaces – some quantification
4. Possible directions for the future

1. What is this thing actually, ‘Drug Discovery’?

- Drug *discovery* implies:
  ‘We don’t have enough!
  We shall set our sails
  and we need *more*, we
  need to… *discover!’*

- But hypotheses are
  (always were!) cheap
  and plentiful

- Drug discovery is instead
  about *de-risking, and*
  retaining what works

Image Source: Recursion
Pharmaceuticals' S-1 SEC Filing

1. What is this thing actually, ‘Drug Discovery’?

- Drug *discovery* implies: ‘We don’t have enough! We shall set our sails and we need *more*, we need to… *discover!’*

- But hypotheses are (always were!) cheap and plentiful

- Drug discovery is instead about *de-risking, and retaining what works*

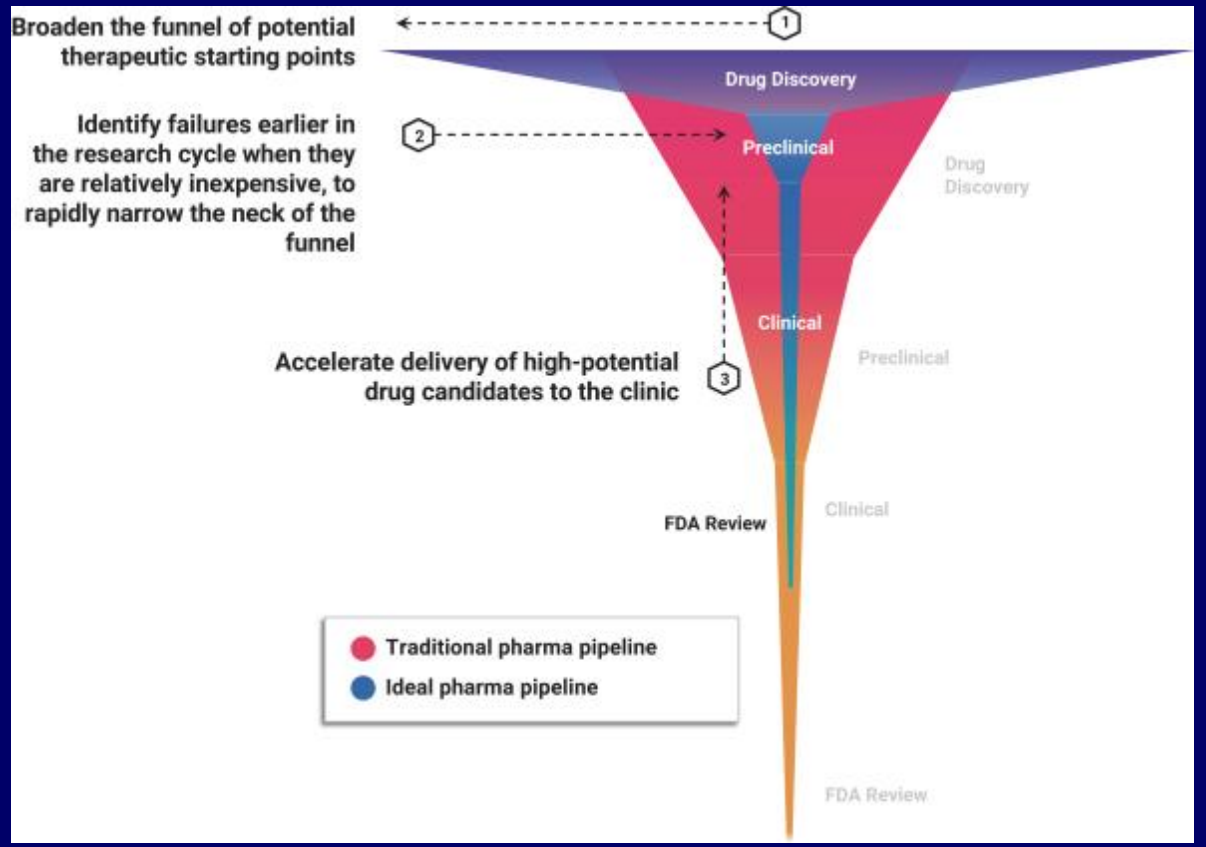


Image Source: Recursion Pharmaceuticals' S-1 SEC Filing

1. What is this thing actually, ‘Drug Discovery’?

- Drug *discovery* implies: ‘We don’t have enough! We shall set our sails and we need *more*, we need to… *discover!’*

- But hypotheses are (always were!) cheap and plentiful

- Drug discovery is instead about *de-risking, and retaining what works*

The point is *not* to stick in *more* up here… *(~‘increase recall’)*

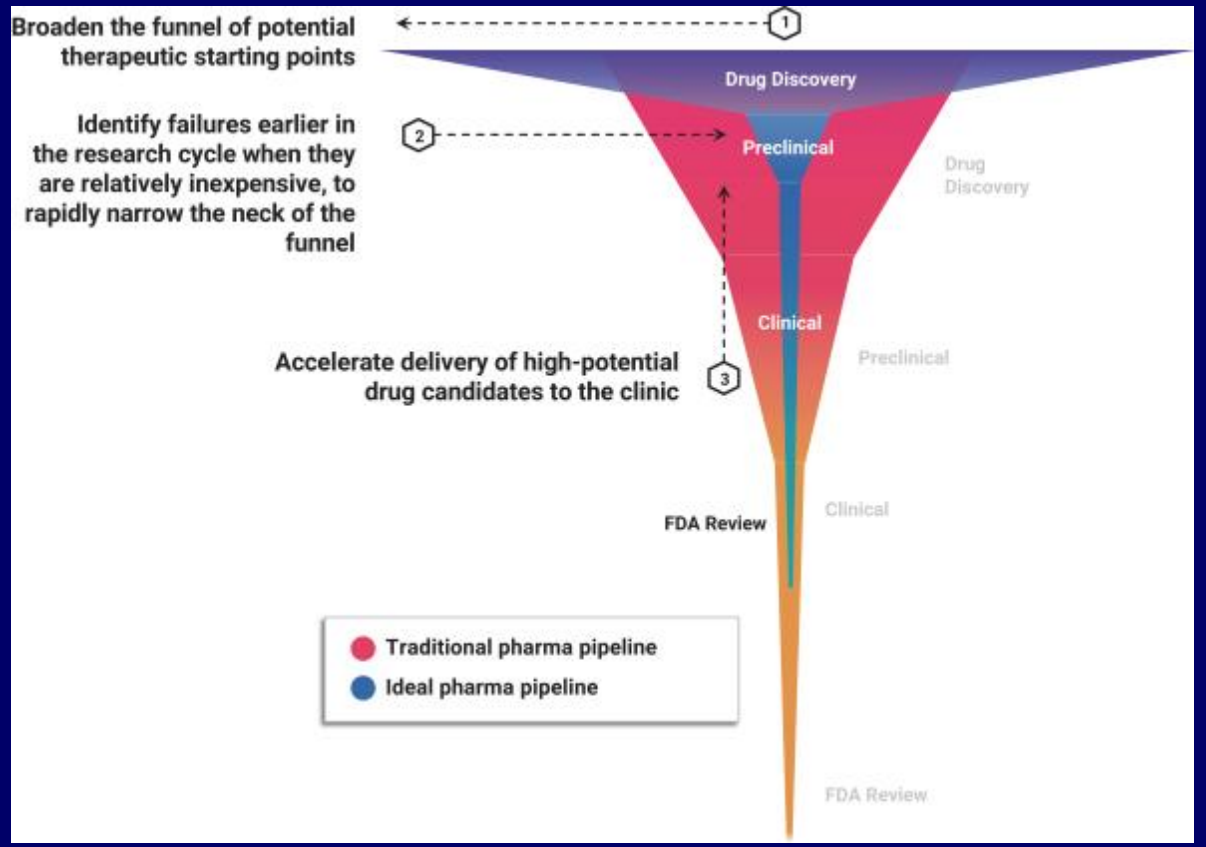


Image Source: Recursion Pharmaceuticals' S-1 SEC Filing

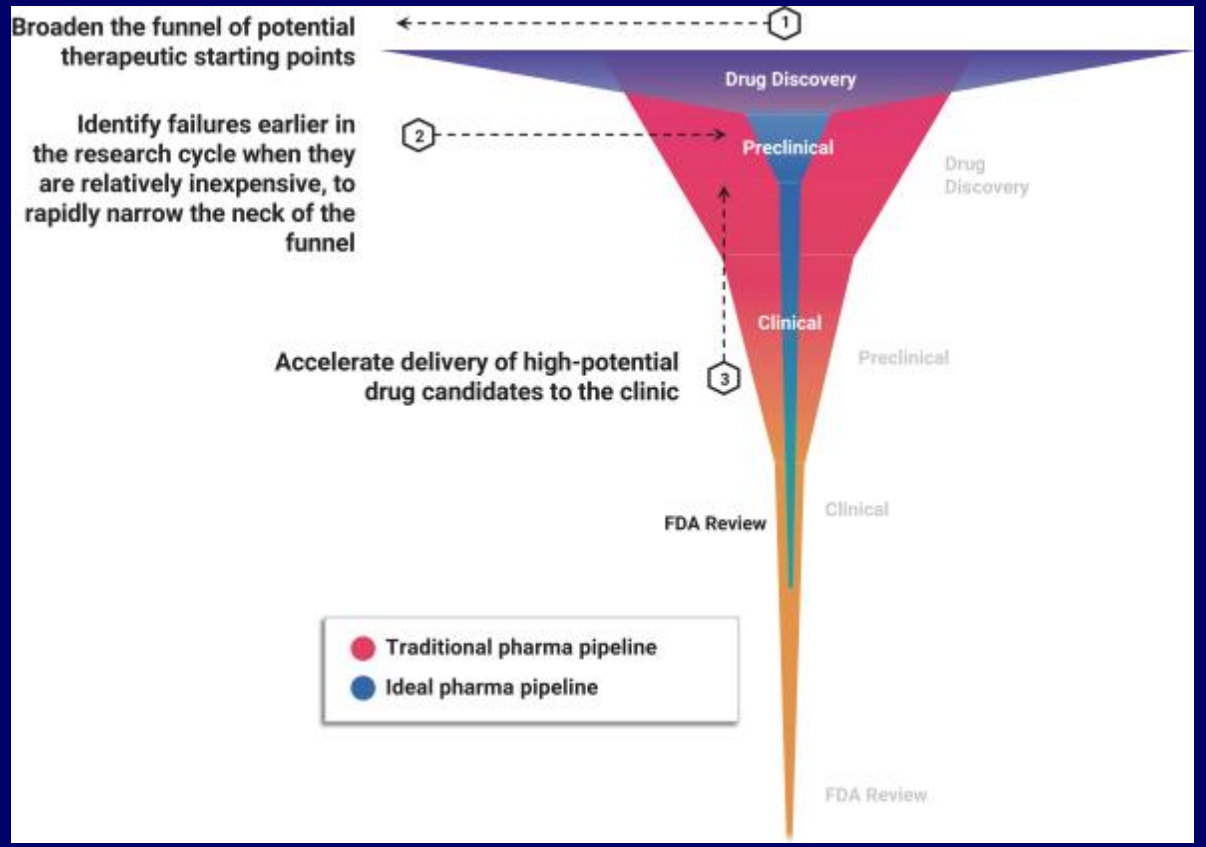
1. What is this thing actually, ‘Drug Discovery’?

- Drug *discovery* implies: ‘We don’t have enough! We shall set our sails and we need *more*, we need to… *discover!’*

- But hypotheses are (always were!) cheap and plentiful

- Drug discovery is instead about *de-risking, and retaining what works*

The point is *not* to stick in *more* up here… *(~‘increase recall’)*



The point is *to have more come out down there (~‘increase precision’)*

Image Source: Recursion Pharmaceuticals' S-1 SEC Filing

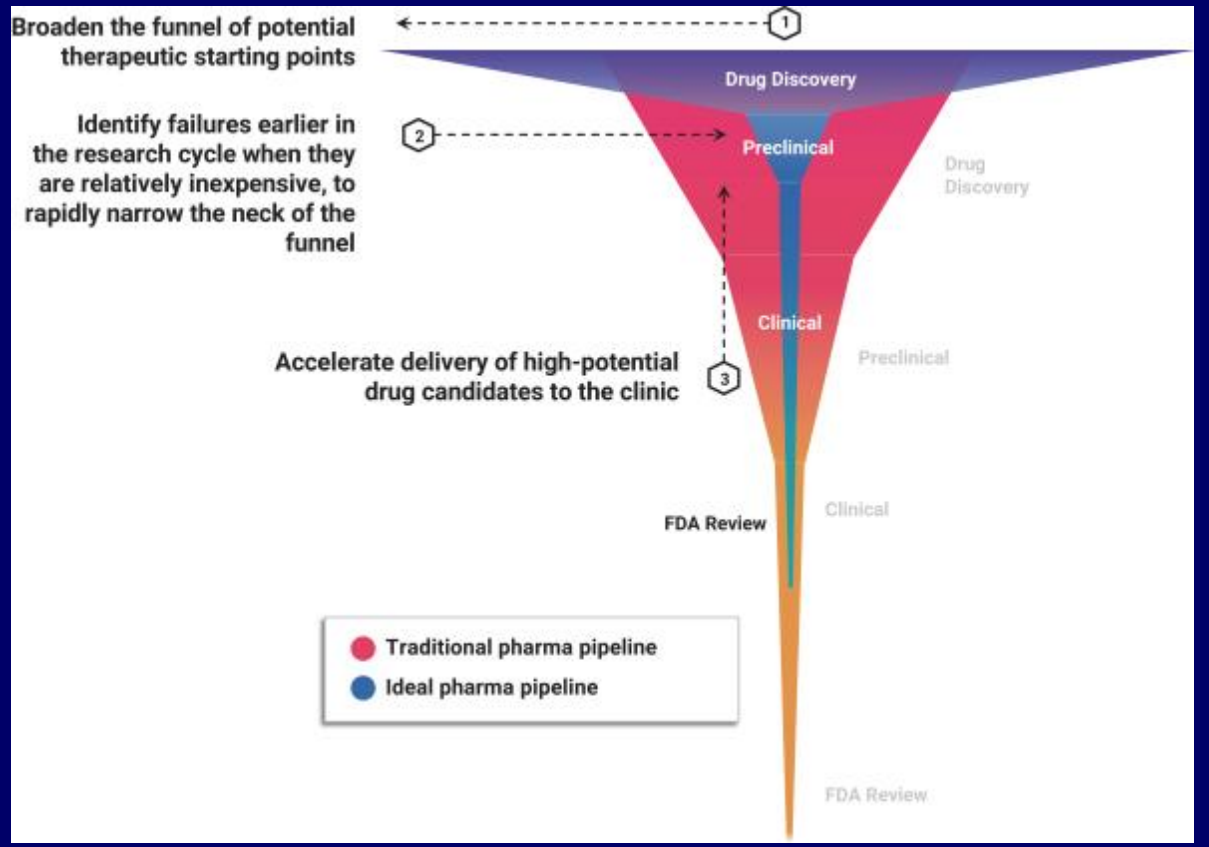
1. What is this thing actually, ‘Drug Discovery’?

- Drug *discovery* implies: ‘We don’t have enough! We shall set our sails and we need *more*, we need to… *discover!’*

- But hypotheses are (always were!) cheap and plentiful

- Drug discovery is instead about *de-risking, and retaining what works*

The point is *not* to stick in *more* up here… *(~‘increase recall’)*



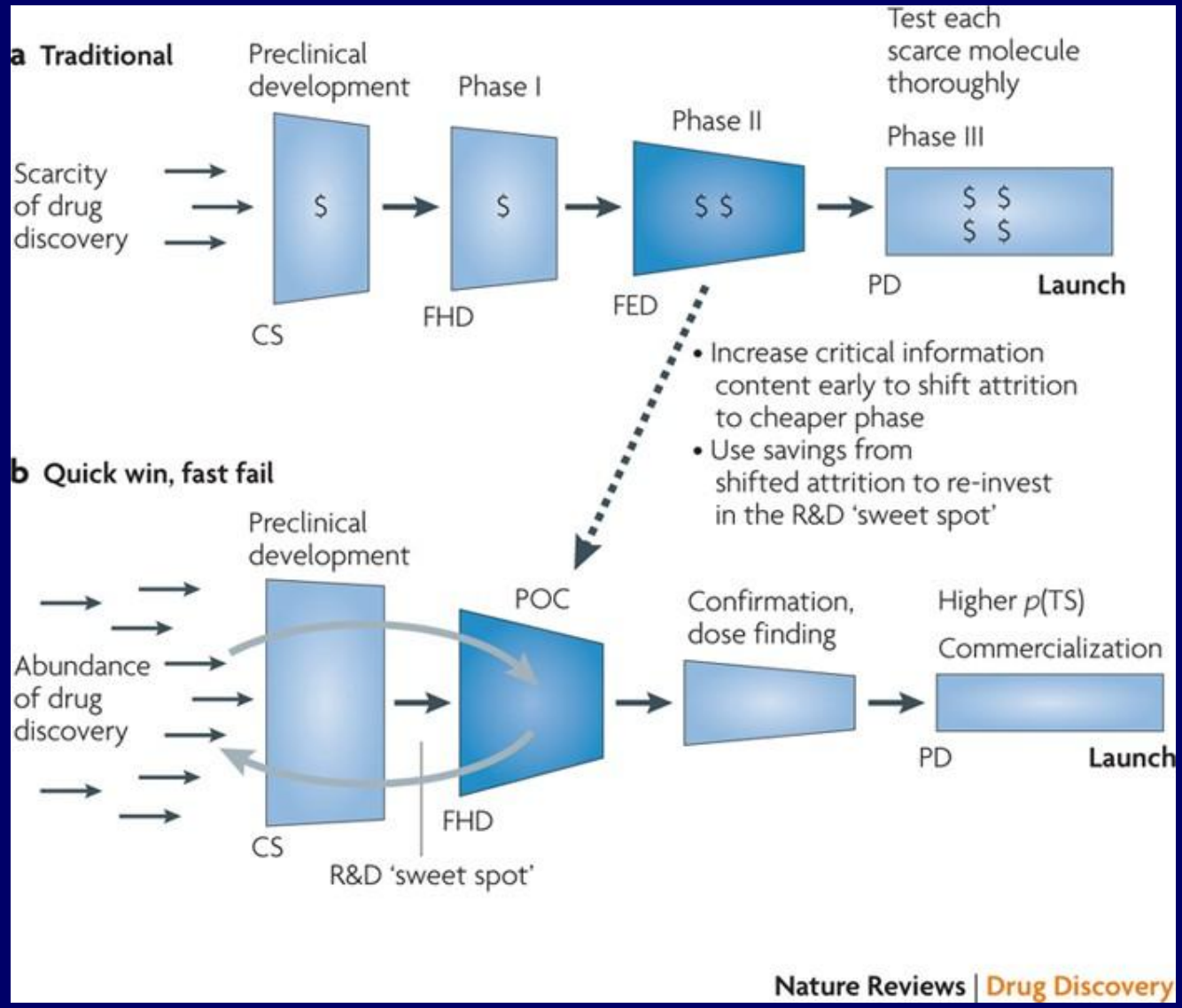
… by having the right hypotheses and selection criteria in place

‘More discovery’ is of limited use as we will see

The point is *to have more come out down there (~‘increase precision’)*

Image Source: Recursion Pharmaceuticals' S-1 SEC Filing

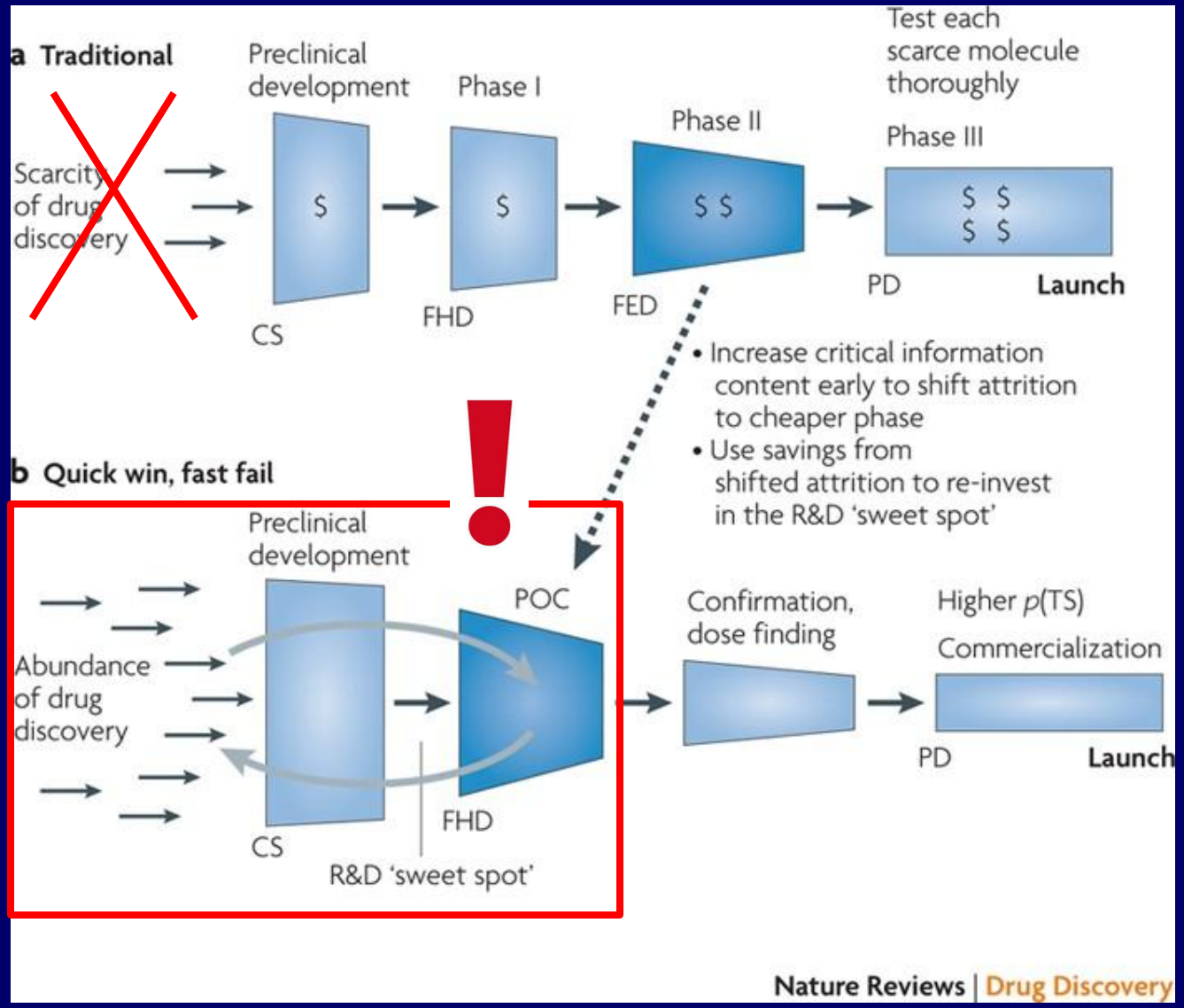
Drug discovery is… sifting through plenty of options, in data-poor spaces



Paul et al., How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery 2010, 9, pages 203–214 (2010)

Drug discovery is… sifting through plenty of options, in data-poor spaces

- We live in an *abundance of drug discovery* world!
- We don't need to '*discover*', we need to *de-risk*!

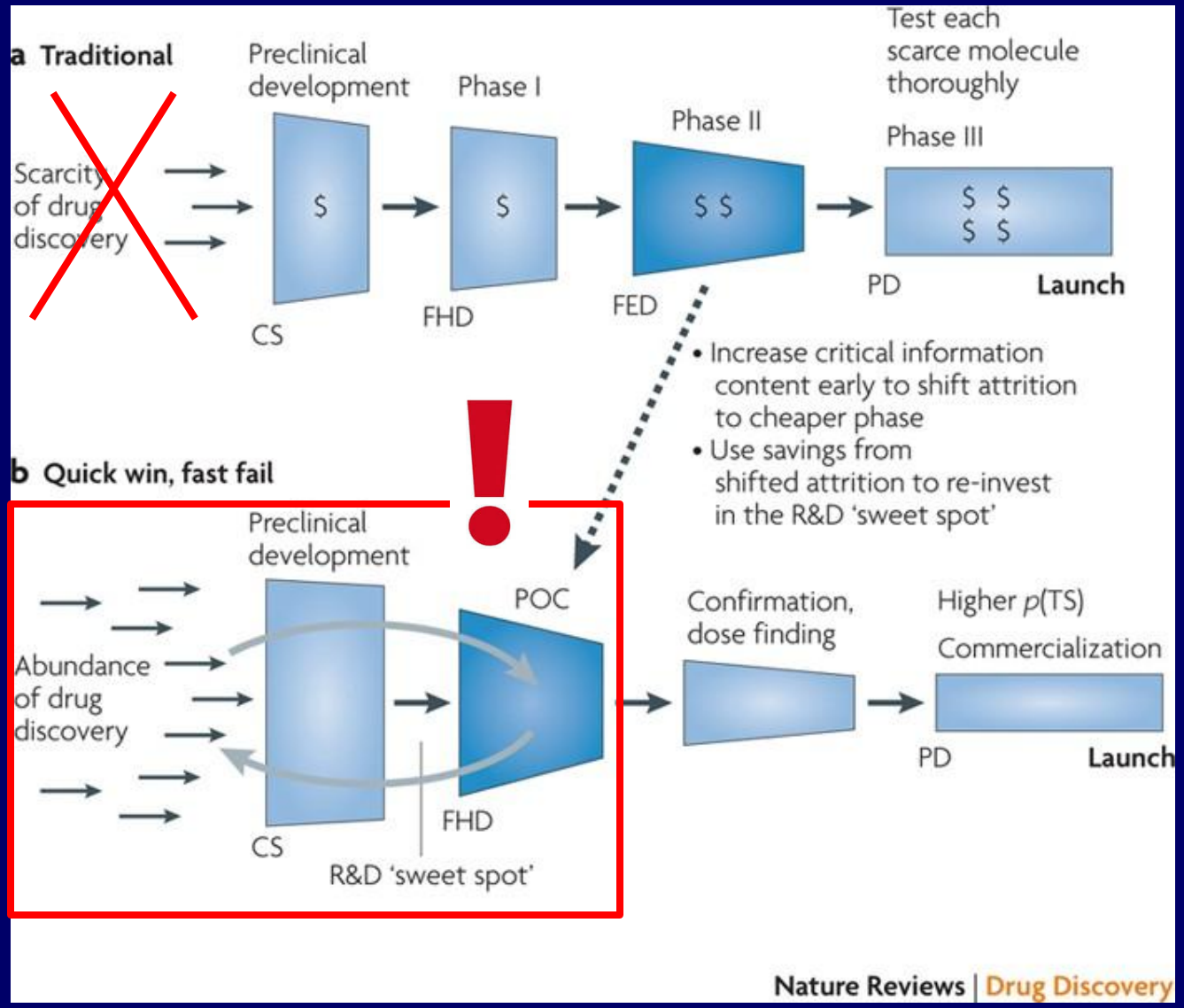


Paul et al., How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery 2010, 9, pages 203–214 (2010)

Drug discovery is… sifting through plenty of options, in data-poor spaces

- We live in an *abundance of drug discovery* world!
- We don't need to '*discover*', we need to *de-risk*!

- As some say: 'Fail fast, fail cheap'
- Personal preference: 'Kill fast, kill early, and kill completely'
(Novartis postdoc seminar, MA, 2008)



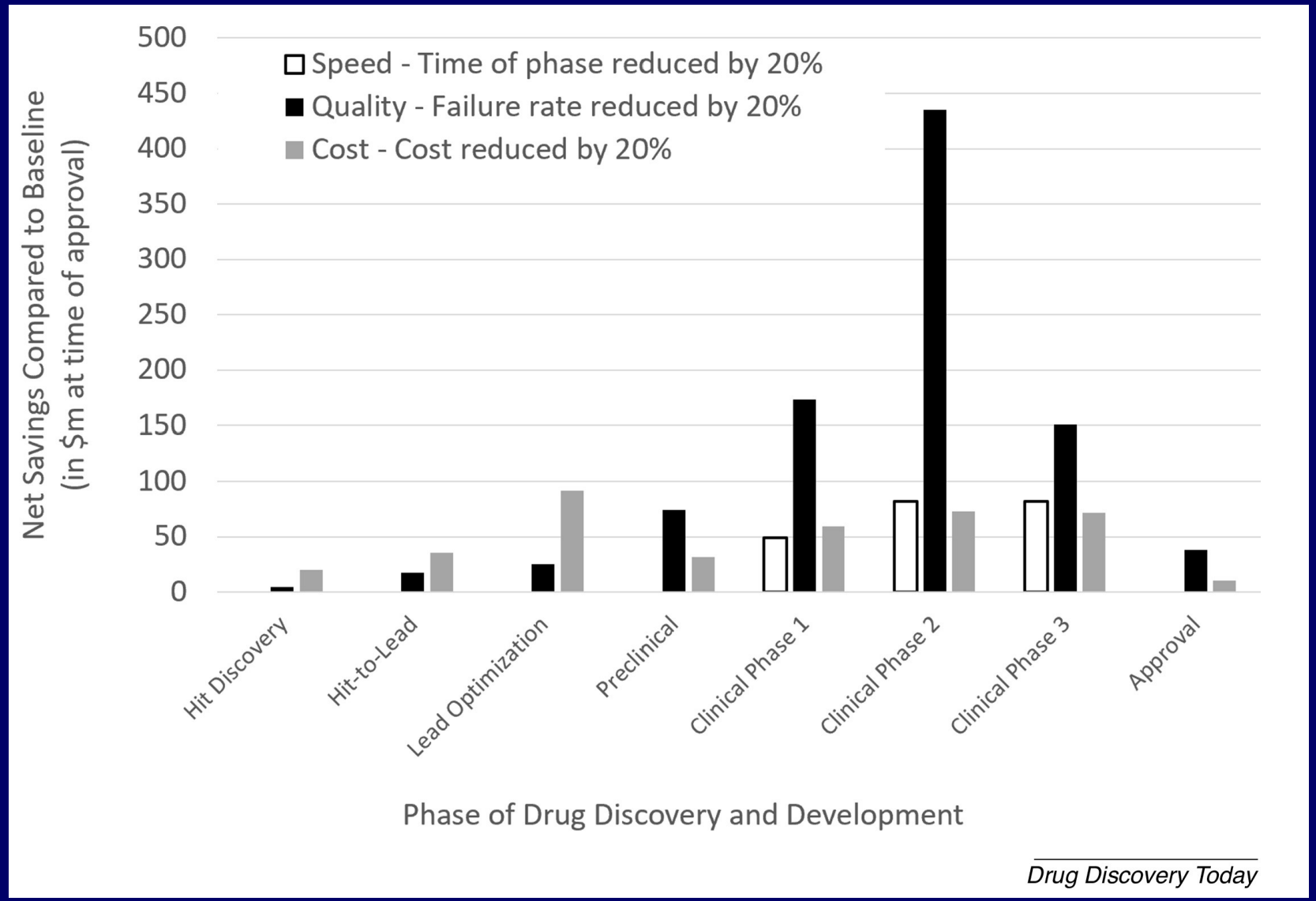
Paul et al., How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery 2010, 9, pages 203–214 (2010)

THE GOAL: Clinically relevant decisions, related to (mostly) efficacy and (also) safety

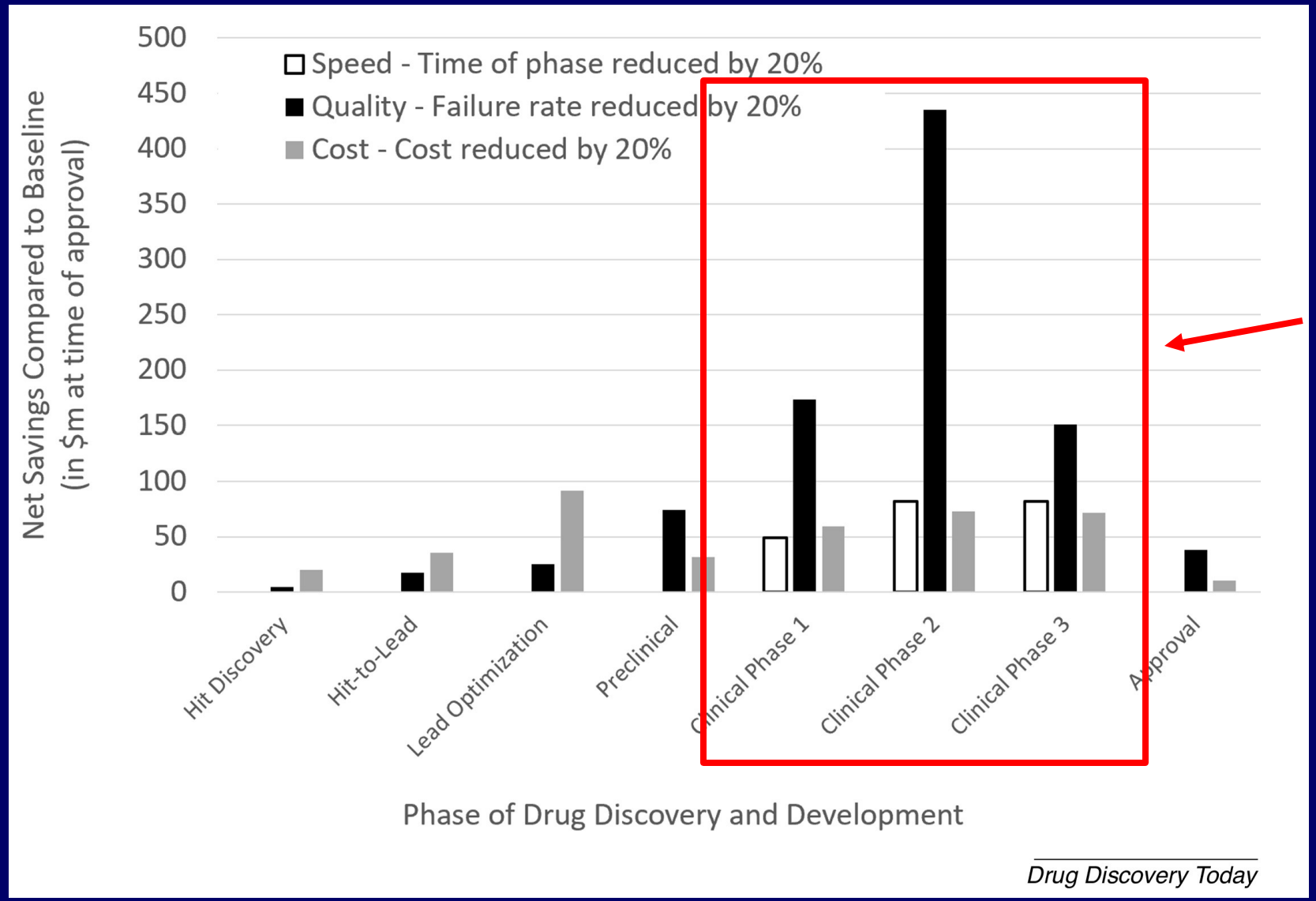
Bender and Cortes, Drug Discovery Today 2021

Bender et al. NRDD 2026

Morgan et al. NRDD 2018



THE GOAL: Clinically relevant decisions, related to (mostly) efficacy and (also) safety



What matters is **the right** compound in the right patient (dosed in the right way) – *i.e.* clinically relevant *precision*

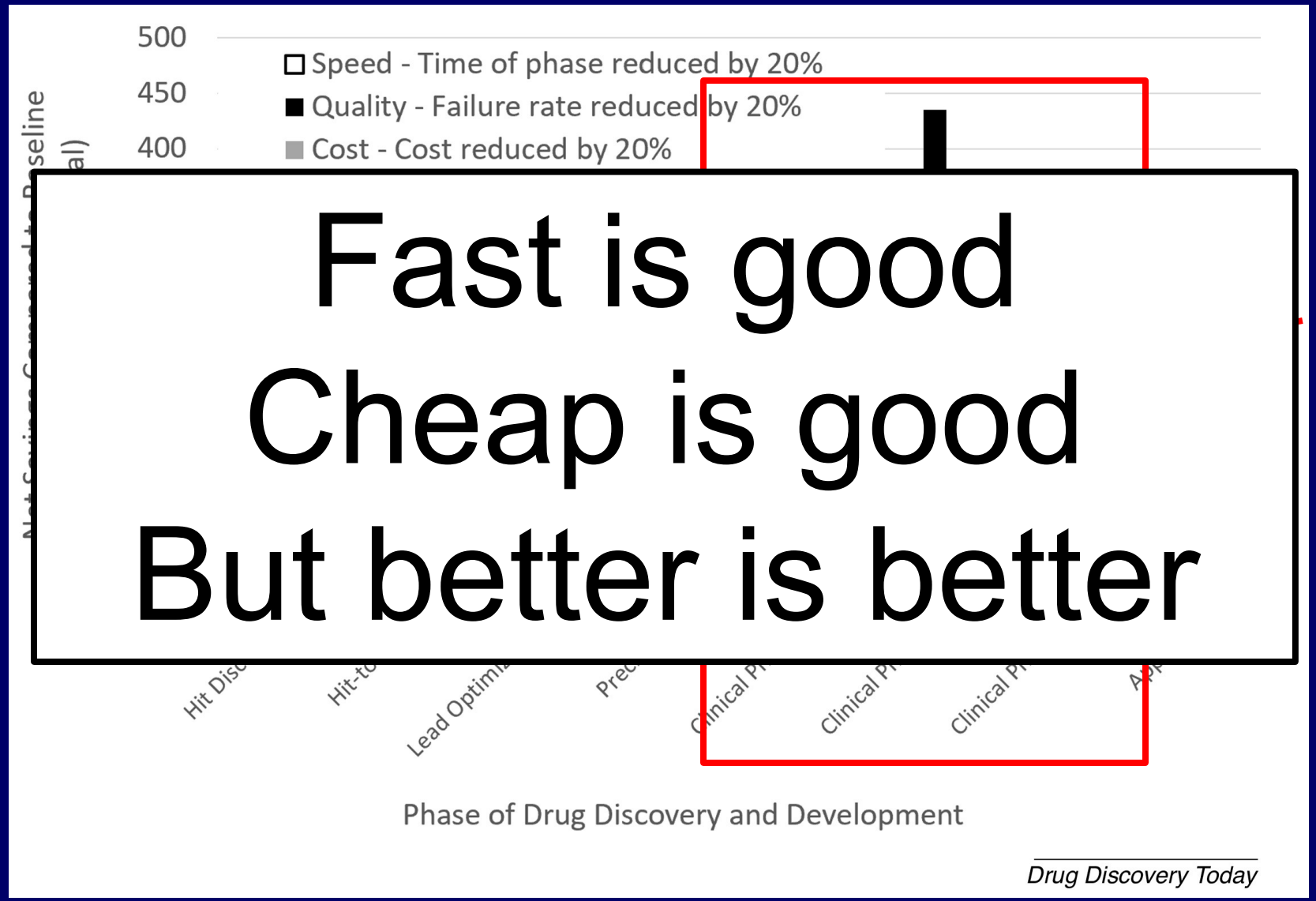
Everything else matters less. *Far* less.

Bender and Cortes, Drug Discovery Today 2021

Bender et al. NRDD 2026

Morgan et al. NRDD 2018

THE GOAL: Clinically relevant decisions, related to (mostly) efficacy and (also) safety



Bender and Cortes, Drug Discovery Today 2021

Bender et al. NRDD 2026

Morgan et al. NRDD 2018

What matters is **the right** compound in the right patient (dosed in the right way) – *i.e.* clinically relevant *precision*

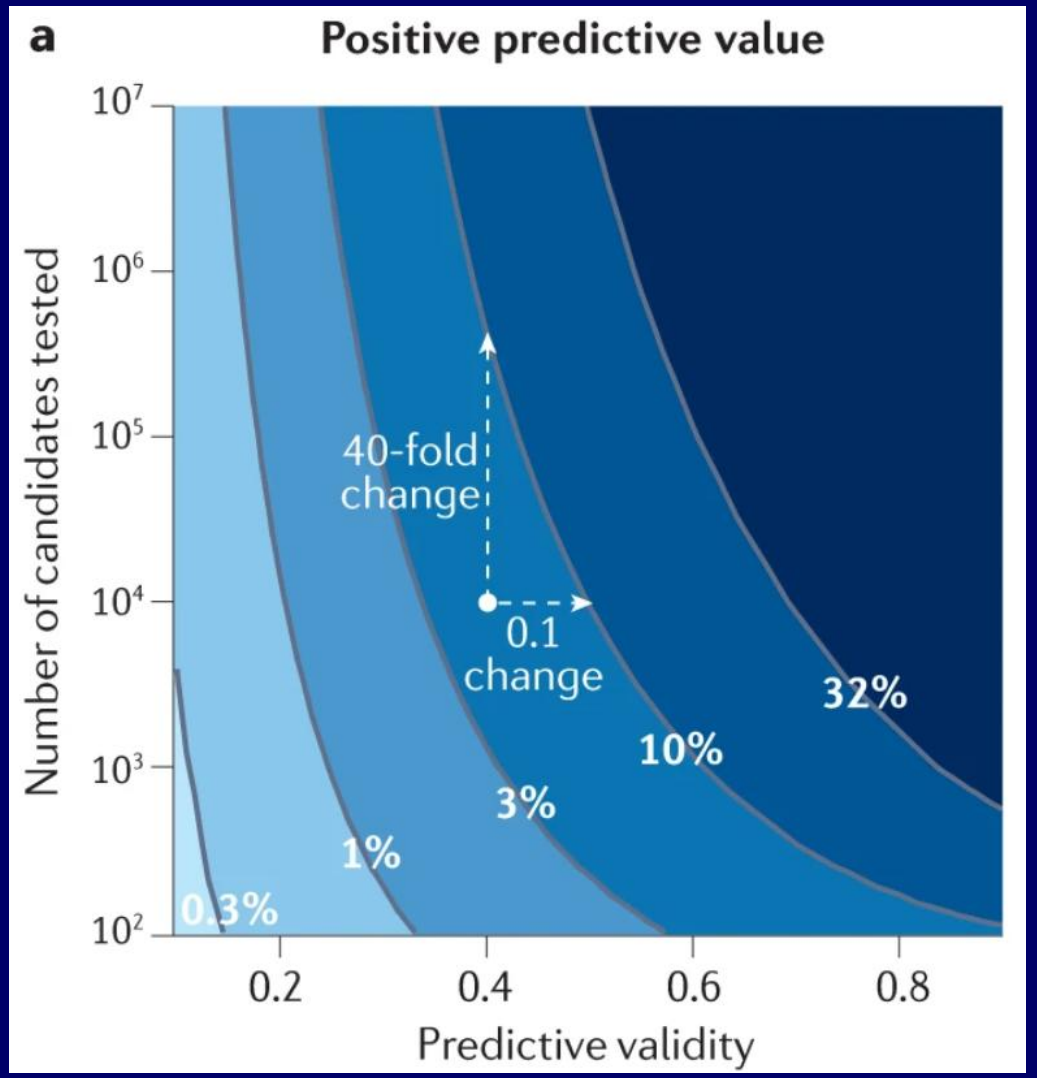
Everything else matters less. *Far* less.

A 10% better predictive validity is worth ca 10-40x the number of compounds tested (!)

Jack and I blow into the same trumpet:

"We need better, not *more (of not good)*"

0.1 better predictive value is worth ~*40x* the throughput (!!!)



Scannell *et al.* **Predictive validity in drug discovery: what it is, why it matters and how to improve it.** Nature Reviews Drug Discovery 2022

But the *slight* problem is…

- We need to do this de-risking in
  - **High-dimensional (‘100-1000s’ of dimensions)**
  - **Rugged (local, e.g. chemical series-dependent)**
  - **Data-poor (sometimes only 100s of data points, esp. with in vivo relevance)**
  - **Meta-data (context)-incomplete (e.g. assay annotations, genomics, …)**

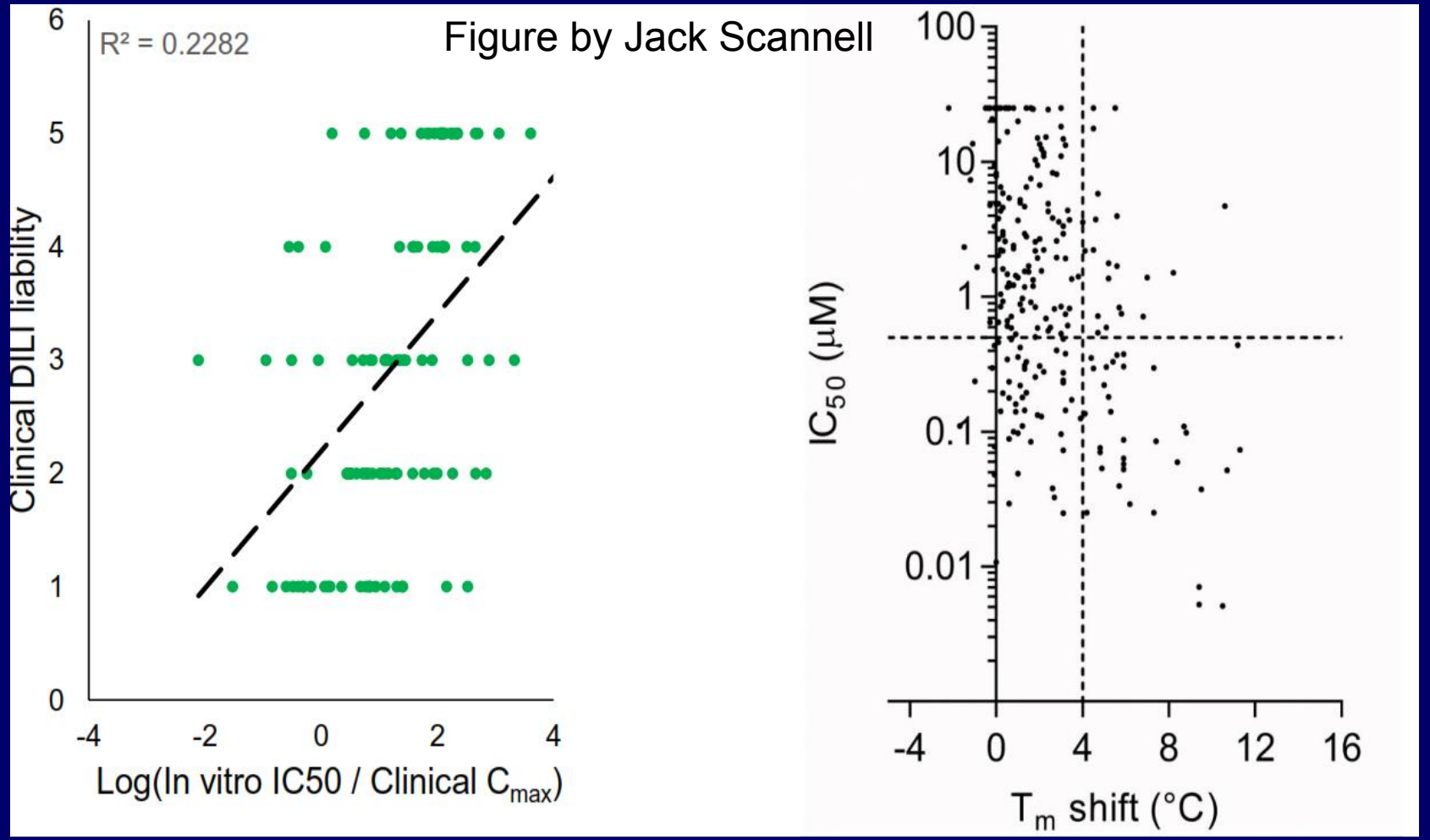
decision spaces, which also generally have

  - **Low predictivity for *human translation*!**

We need to take care of (selected list): On-targets, Off-targets, Bioavailability, Solubility, Cmax, Half life, Microsomal stability, Accumulation in organelles, Accumulation in organs, Permeability, Target engagement, Organ Toxicity (Major Organs: Liver, Kidney, Lung, CNS, Cardiovascular) plus other ‘minor’ organs…

… and we (often) don't really have 'good' (predictive) data

Figure by Jack Scannell



Left: Clinical DILI liability related to Cmax-corrected organoid-derived IC50 values, with low correlation between both values (lower liability index values indicate higher clinical liability)

Right: Low correlation of enzymatic and thermal-shift derived activity data.

Solely feeding such data with low predictivity into 'AI' models will not lead to better individual decisions, and hence clinical outcomes.

Proctor WR *et al.*. Utility of spherical human liver microtissues for prediction of clinical drug-induced liver injury. Arch Toxicol. 2017 Aug;91(8):2849-2863.

Rudolf AF *et al.*. A comparison of protein kinases inhibitor screening methods using both enzymatic activity and binding affinity determination. PLoS One. 2014 Jun 10;9(6):e98800.

Similar low predictivity of *in vivo* readouts – e.g. toxicity

Positive toxic event concordance of toxicity is ‘quite good’:

‘The results showed the true positive Human Toxicity concordance rate of 71% for rodent and nonrodent species, with nonrodents alone being predictive for 63% of Human Toxicities and rodents alone for 43%.’

Regulatory Toxicology and Pharmacology. 2000;32(1):56-67

… BUT…

‘Our finding that absence of animal toxicity is of limited predictivity for the lack of adverse events in humans was already reported in previous studies’

Regulatory Toxicology and Pharmacology. 2018;96:94-105

For other examples (related to efficacy) see e.g. Perel, P. et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 334, 197 (2007) & https://humanspecificresearch.org/the-reality-of-drug-discovery-and-development

Is *bigger*, is *more* better (in terms of data)?

- How much data do I want? MBs? GBs? TBs? PBs?

- Single-cell and time resolved and multimodal and multiomics…?

Is *bigger*, is *more* better (in terms of data)?

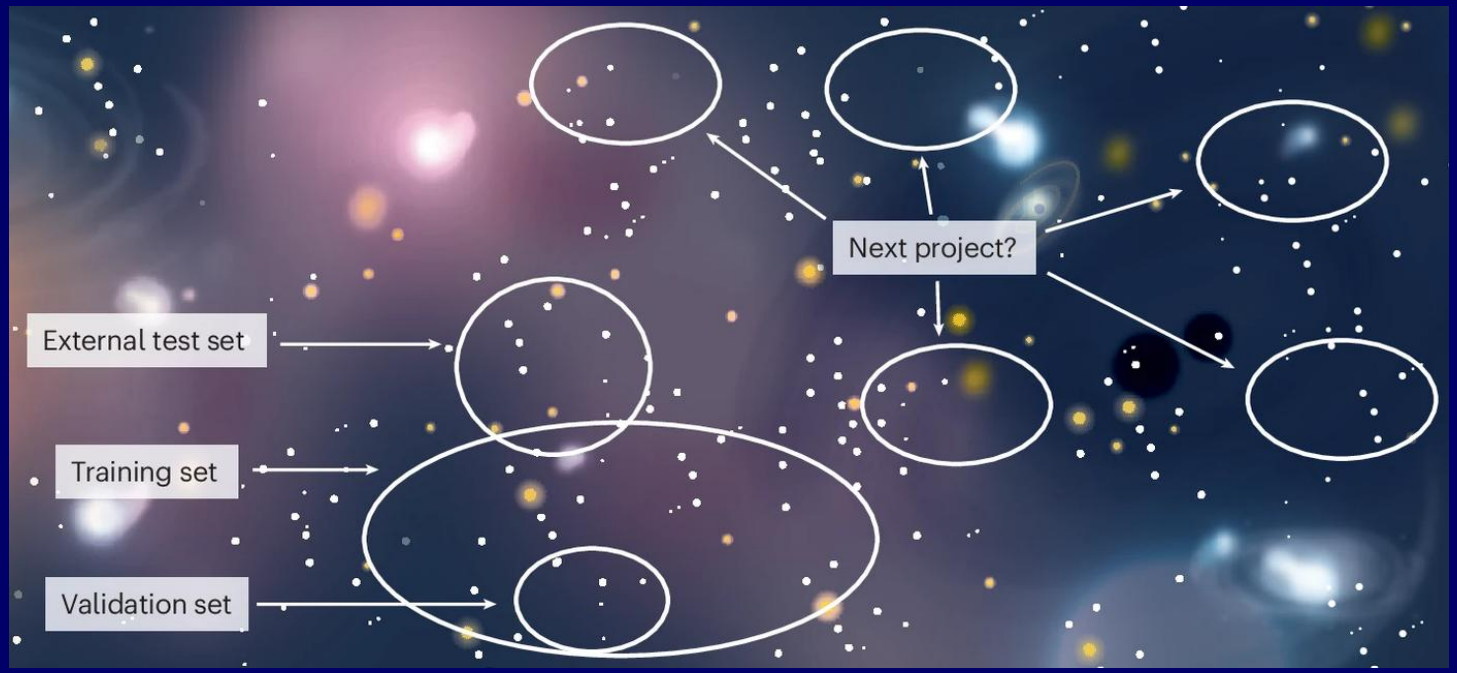
- How much data do I want? MBs? GBs? TBs? PBs?
- Single-cell and time resolved and multimodal and multiomics…?
- I want precisely *1 bit of data* (per decision that needs to be made)
- But *the right bit* – the one that tells me, ‘yes or no’ (is this the right molecule, for the given purpose, etc)

Bottom line

- Drug discovery is not about ‘finding many shiny new things’, it is about *de-risking and clinical translation*

- *Take home: Quality* (= mostly *precision*, due to ‘abundance’ setting) of *translationally relevant decisions* is what matters in drug discovery

2. Many current models do not address this need: Often ‘validation’ is just an exercise to generate numbers with little relevance – e.g.



Difficulties with extrapolation are intrinsic to the problem we face

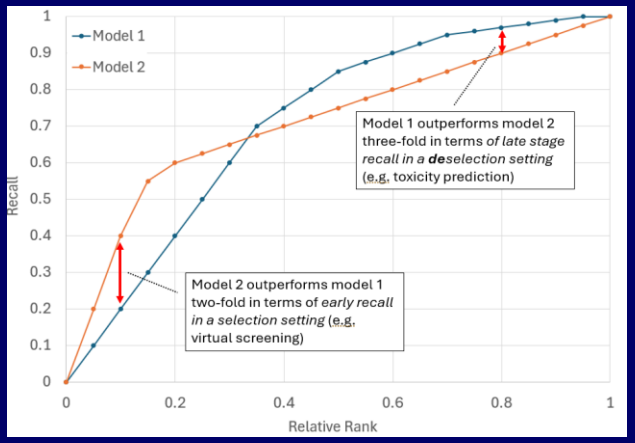
Some other problems (metrics etc) are what we can influence ourselves with better decisions

- **Future projects will *by definition* be outside the training set distribution**
- In particular key due to *highly local nature* of chemical space for directional properties
- Also time-split doesn’t help, it’s just different galaxies which equally don’t matter

Performance measured retrospectively will not hold prospectively (in future projects)

Bender *et al.* Artificial intelligence in drug discovery — what it is, where we stand and the path forward. *Nature Reviews Drug Discovery* 2026 (25) 855–876.

Global metrics are virtually never predictive of real-world performance, use-case metrics are needed



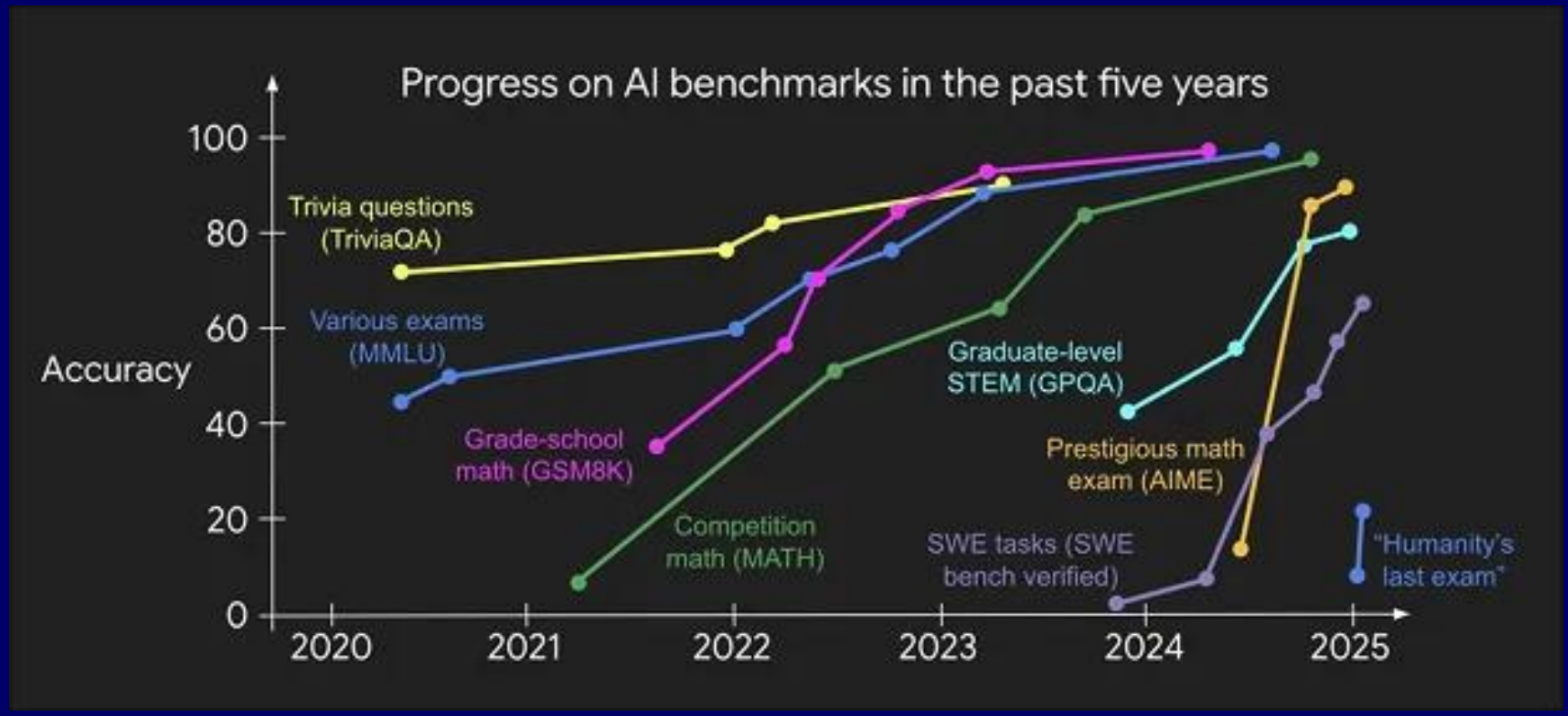
- **Point-error based metrics** (RMSE etc): They sound p*retty good*… if on average there is no change anyway, and that's what you just always predict (but it's not what you are interested in!)
- **Correlation-based metrics** (R2 etc): Often high global values … but in decision-making situations you are hardly ever interested in overall distributions (you need within-group differentiation, etc)
- **Distribution-based metrics** (AUC etc) do not discriminate between early and late recall settings etc.
- **Use-case relevant metrics are better choice** (e.g. 'recall of compounds with clearance > liver blood flow' for in vivo PK models)
- Applies also to many other models of course (non-goal directed metrics in early generative chemical models, etc)

Conclusion: We produce many numbers, but they don't mean all that much

Benchmarks, but…

We are IMO in an ‘Illusion of progress’ situation *(with respect to overall outcome)* …

Another benchmark saturated:



https://agents-last-exam.org/blogs/why-agents-last-exam



Recommended album: Staind, ‘The Illusion of Progress’

But when translating to ‘the real world’ there are *very few* prospective, translationally relevant applications

Example from current literature and level of ‘validation’ performed

Accelerating scientific breakthroughs with an AI co-scientist

‘Drug repurposing for acute myeloid leukaemia’

- Biminetinib found to have 7nM IC50 in AML cell lines
- … but putting a ***known*** **cytotoxic kinase inhibitor** with hundreds of data points against kinases *into a cell line* **and finding it to *be cytotoxic* is no ‘repurposing’**
- Current validations and use cases involving *decision making* are very poor and have little clinical relevance

Other aspects of ‘validation’ to be careful with (selected examples)

- **‘Retrospective validation’**

  Read as: ‘*My model can predict yesterday’s weather*’
- **‘Our agent performs better than (vanilla) LLM X’**

  Read as: ‘We consciously neglected state-of-the-art (domain-specific) solutions and used an artificially degraded baseline’
- **No full performance metrics presented (precision and recall etc):**

  Read as: ‘We have no idea if the model worked, we have hand-picked the presented examples anyway’

Etc.

- The proof of the pudding is in the eating – not ‘validation’, but *application*
- If academia/a small start-up can’t go the whole way there are reasons, well-capitalized companies don’t have any of those excuses

3. Data and models vs decision spaces and process impact

3. Data and models vs decision spaces and process impact

- Years ago I discovered the ‘power of literature validation’, say linking a predicted target to a disease
  - The beauty of it: *It always works*

3. Data and models vs decision spaces and process impact

- Years ago I discovered the ‘power of literature validation’, say linking a predicted target to a disease
  - The beauty of it: *It always works*
- Experience using e.g. knowledge graphs for compound selection/repurposing
  - 2 or 3 hops, and you are everywhere in your network, entirely unable to prioritize (globally; locally you sometimes get more lucky)

3. Data and models vs decision spaces and process impact

- Years ago I discovered the ‘power of literature validation’, say linking a predicted target to a disease
  - The beauty of it: *It always works*
- Experience using e.g. knowledge graphs for compound selection/repurposing
  - 2 or 3 hops, and you are everywhere in your network, entirely unable to prioritize (globally; locally you sometimes get more lucky)
- Everyone happily waffles (LLMs, people, irreproducible data, …)
  - There is just more and more stuff to wade through

3. Data and models vs decision spaces and process impact

- Years ago I discovered the ‘power of literature validation’, say linking a predicted target to a disease
  - The beauty of it: *It always works*
- Experience using e.g. knowledge graphs for compound selection/repurposing
  - 2 or 3 hops, and you are everywhere in your network, entirely unable to prioritize (globally; locally you sometimes get more lucky)
- Everyone happily waffles (LLMs, people, irreproducible data, …)
  - There is just more and more stuff to wade through

The common problem that emerges: *Precision*!

Let's look at some numbers – e.g. hit finding

- 500k out of /170m library are likely 'hits' against D4 (estimation); ~0.3%

  Ultra-large library docking for discovering new chemotypes

  https://www.nature.com/articles/s41586-019-0917-9

- 0.1% hit rate (up to ~1%) in HTS, biased library

  Changing the HTS Paradigm: AI-Driven Iterative Screening for Hit Finding

  https://pmc.ncbi.nlm.nih.gov/articles/PMC7838329/


Article | Published: 31 May 2009

Quantifying biogenic bias in screening libraries

Jérôme Hert, John J Irwin, Christian Laggner, Michael J Keiser & Brian K Shoichet

Nature Chemical Biology **5**, 479–483 (2009) | Cite this article

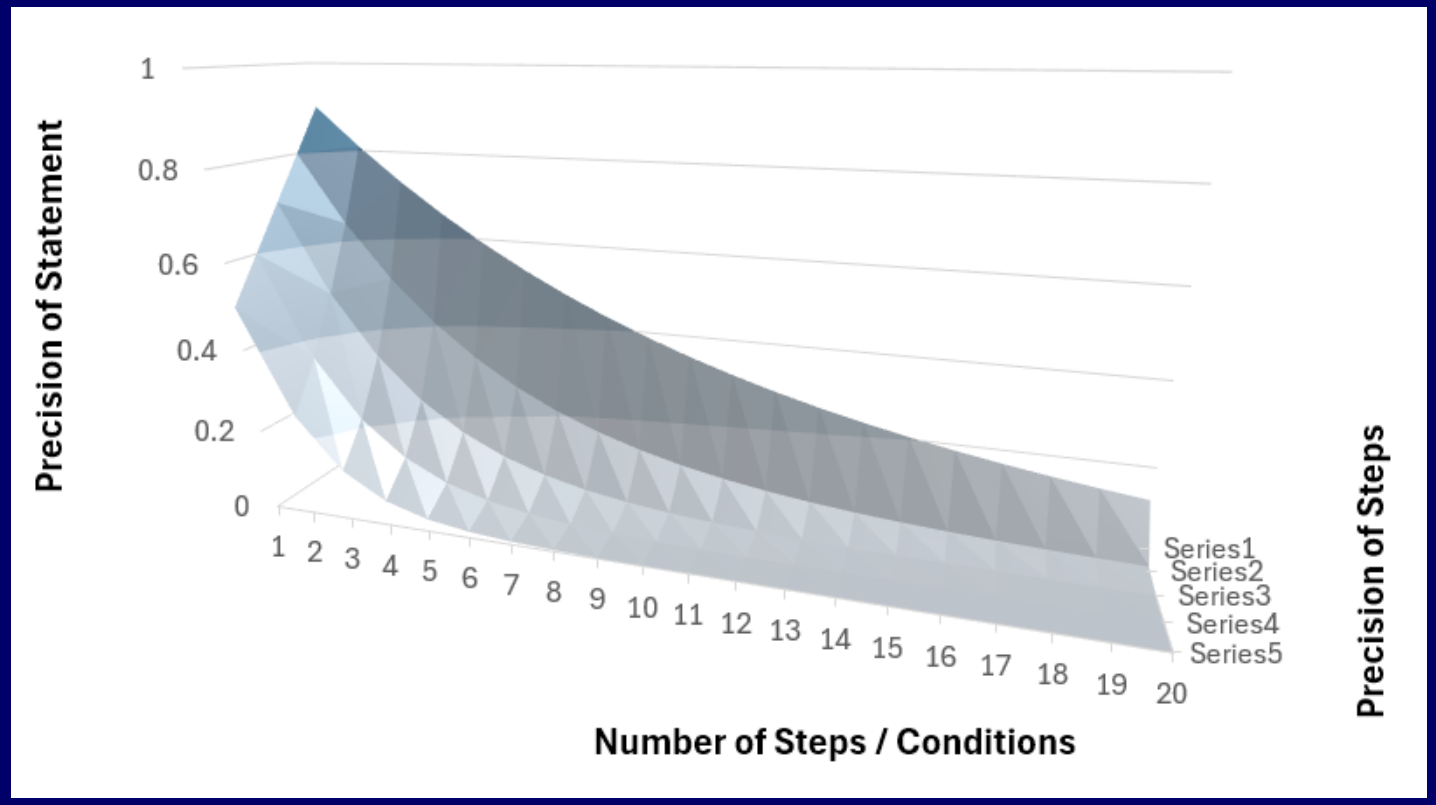
- Why does it work? Large space (~10^{60}), but *biased libraries*, *local precision* of model gives reasonable hit rate (as opposed to *global* model used in an *unbiased* library requiring *high accuracy*)
- **Helpful biases** we can exploit in screening are in the order of 10^{dozens} (!!), say similarity to metabolites etc

Biological decision space has (far) fewer *helpful biases*

- Ca 7,000 genetic diseases, approx. 20,000 total
- Say 20,000 genes/targetable entities
- 400m disease-gene pairs
- Pharmacology, dose, clinical endpoint, PK, more complex organismal biology, ...

- >10^9 gene-indication-dose-pharmacology combinations
- **Genetics gives only *2.6x* improvement** (Minikel et al. Nature 2024), as opposed to 10^{dozens} (!)
- Far too small helpful bias to get lucky by chance

Quantifying drug discovery: Precision of statement as function of accuracy of steps/conditions

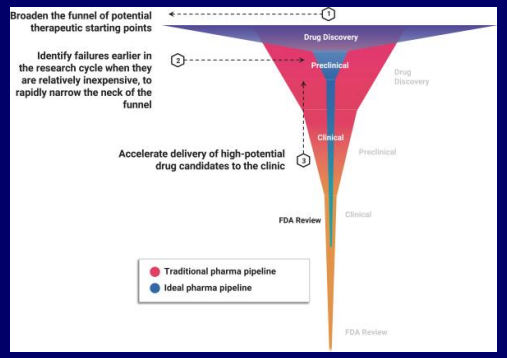


Precision of individual condition	Precision of 20 steps/ conditions
0.9	0.1%
0.8	0.01%
0.7	0.001%
0.6	$4*10^{-5}$
0.5	$9.5*10^{-7}$

- If drug discovery had only 10 conditions (*it has more*) and every condition has a precision of 0.7 overall precision is only 3%!
- Compare to clinical overall technical Probability of Success of 5-15% (Ph1->Approval) - we are doing quite well in this complex space!

Note from the real world: lack of reproducibility of (life science and other) research; lack of negative data

- For de-risking *precision* of statements is key
- Small sample sizes, biases, multiple hypothesis testing etc increase *false positives* in statements/data
- Publication bias *decreases number of true negative statements*
- Estimates:
  - **False positives** ~50% (Ioannidis PLoS Medicine 2005) to 71% (Prinz et al. NRDD 2011)
  - Publication bias ~80% of negative results are not published (Fanelli, PLoS One 2010), **lack of true negatives**



Review Article | Published: 29 September 2026

Reproducibility in biomedical research

Judy Zhong, Debra D'Angelo, Chen Lyu, Ian Xu, Peter K. Enns, Fei Wang & Rainu Kaushal

Nature Medicine **32**, 3564–3577 (2026) | Cite this article

Why Most Published Research Findings Are False

John P. A. Ioannidis

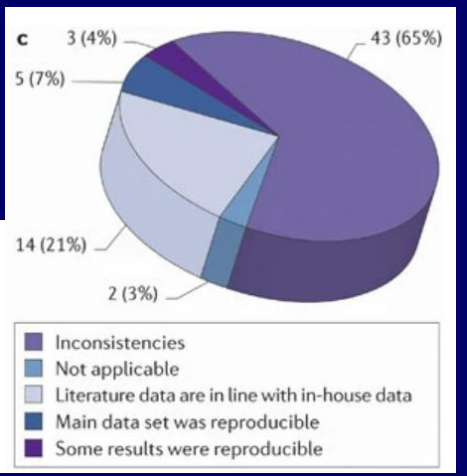
Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124

PLOS Medicine

CAREER NEWS | 22 July 2025

Researchers value null results, but struggle to publish them

Survey finds that fear of reputational harm and a lack of support and publication platforms are among respondents' key concerns.



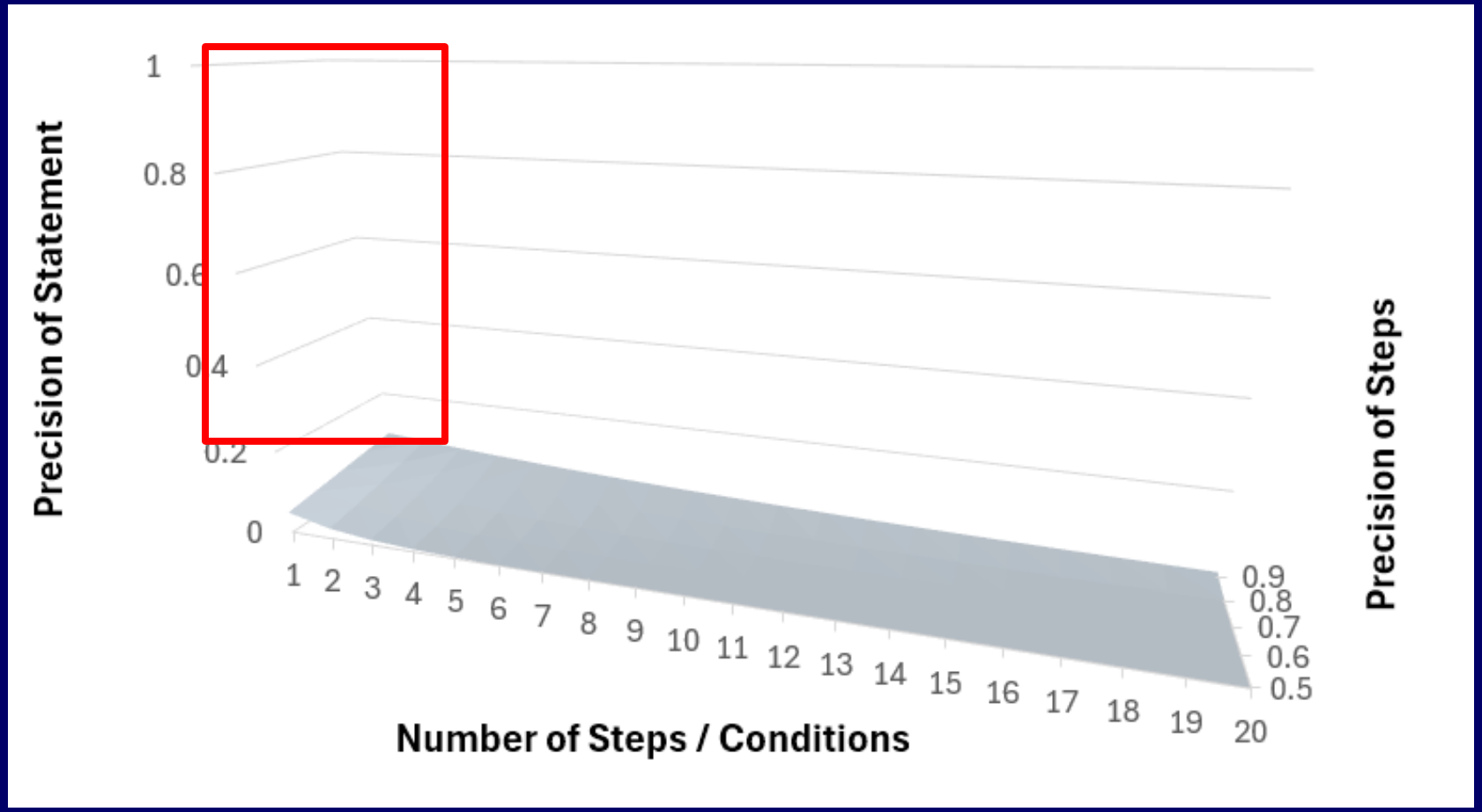
Correspondence | Published: 31 August 2011

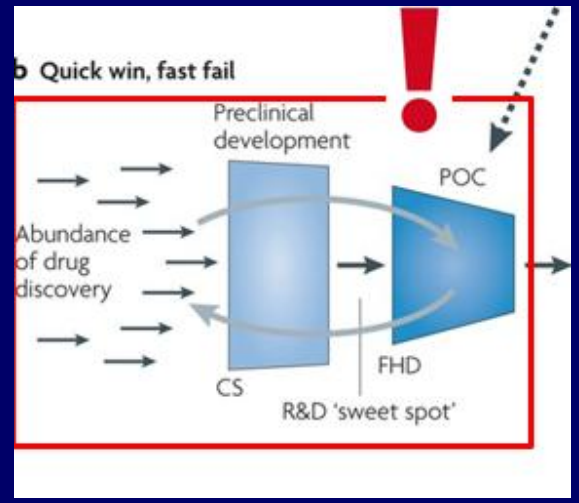
Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange & Khusru Asadullah

Nature Reviews Drug Discovery **10**, 712 (2011) | Cite this article

Taking false positives, lack of true negatives into account… even in low-dimensional spaces precision drops rapidly!





Precision of individual condition	Precision of *10* steps/ conditions
0.09	0.01%
0.08	0.001%
0.07	$8*10^{-5}$
0.06	$4*10^{-6}$
0.05	$9.5*10^{-8}$

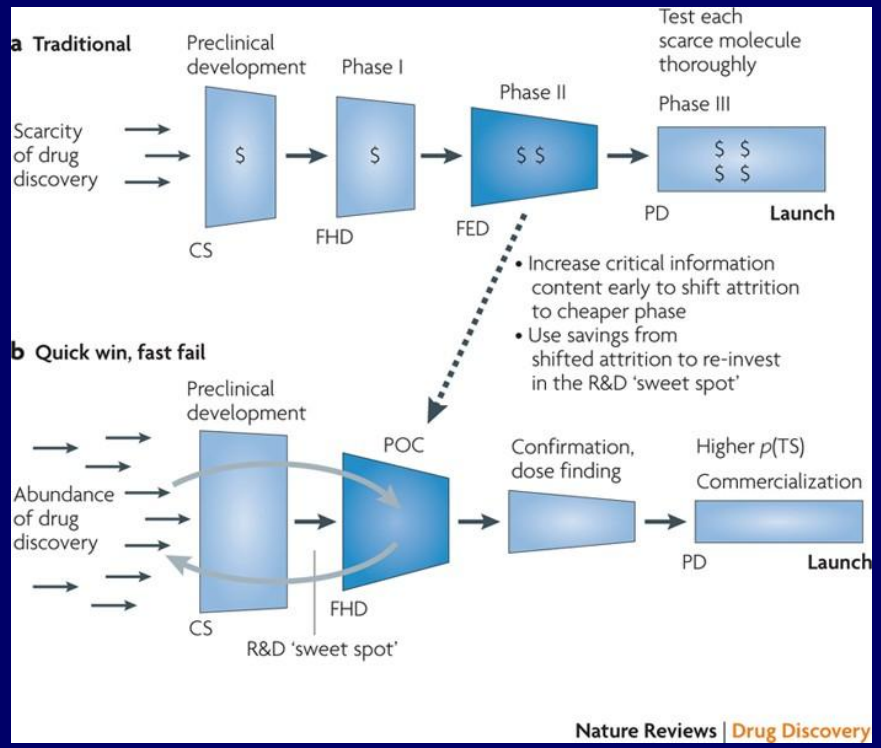
- Even precision of statements involving quite few conditions drops
- Relevant for low-dimensional statements (repurposing etc.)

Bottom line:
Automated decision making will *not* be a piece of cake

We operate in

- **High-dimensional**
- **Rugged (local)**
- **Data-poor**
- **Meta-data (context)-incomplete** *decision spaces*, which also generally have
- **Low translational predictivity**

- We need precision (to de-risk) *BUT*
- We have (apart from the above problems!) too many false positives, too few true negatives to do so
- Leads to lack of precision in de-risking, and arriving at true positive statements

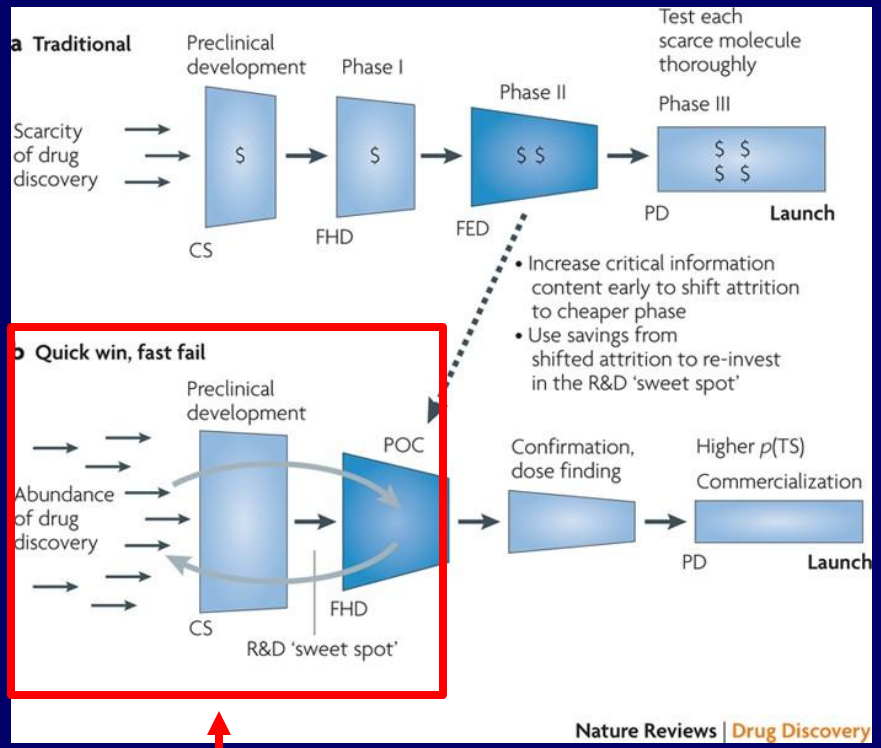


Bottom line:
Automated decision making will *not* be a piece of cake

We operate in

- **High-dimensional**
- **Rugged (local)**
- **Data-poor**
- **Meta-data (context)-incomplete** *decision spaces*, which also generally have
- **Low translational predictivity**

- We need precision (to de-risk) *BUT*
- We have (apart from the above problems!) too many false positives, too few true negatives to do so
- Leads to lack of precision in de-risking, and arriving at true positive statements



Our data often does not allow us to 'fail fast'!

Three types of models in my experience ‘work’, all of which simplify search space (non-exhaustive list)

- Low-dimensional search space, **smooth**, filled with data
  - E.g. logD prediction
- Local searches, supported by heuristics (yes, novelty trade-off, but otherwise we have insufficient precision)
  - Virtual screening – good hit rates in *biased libraries (heuristic)*
  - *Local* precision
- Global, meta-level heuristics (more tricky case since global)
  - E.g. ‘5Rs’ – right drug, patient, tissue, target, dose

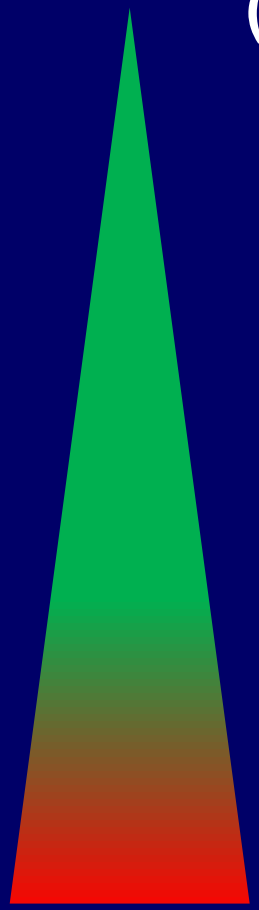
Three types of models in my experience ‘work’, all of which simplify search space (non-exhaustive list)

- Low-dimensional search space, **smooth**, filled with data
  - E.g. logD prediction
- Local searches, supported by heuristics (yes, novelty trade-off, but otherwise we have insufficient precision)
  - Virtual screening – good hit rates in *biased libraries (heuristic)*
  - *Local* precision
- Global, meta-level heuristics (more tricky case since global)
  - E.g. ‘5Rs’ – right drug, patient, tissue, target, dose

Problematic case: High-dimensional, rugged global space with little data and few helpful biases/heuristics… most of life science data!

Will models help in ‘drug discovery’? Depends on data and decision space being dealt with (dimensionality, predictivity)

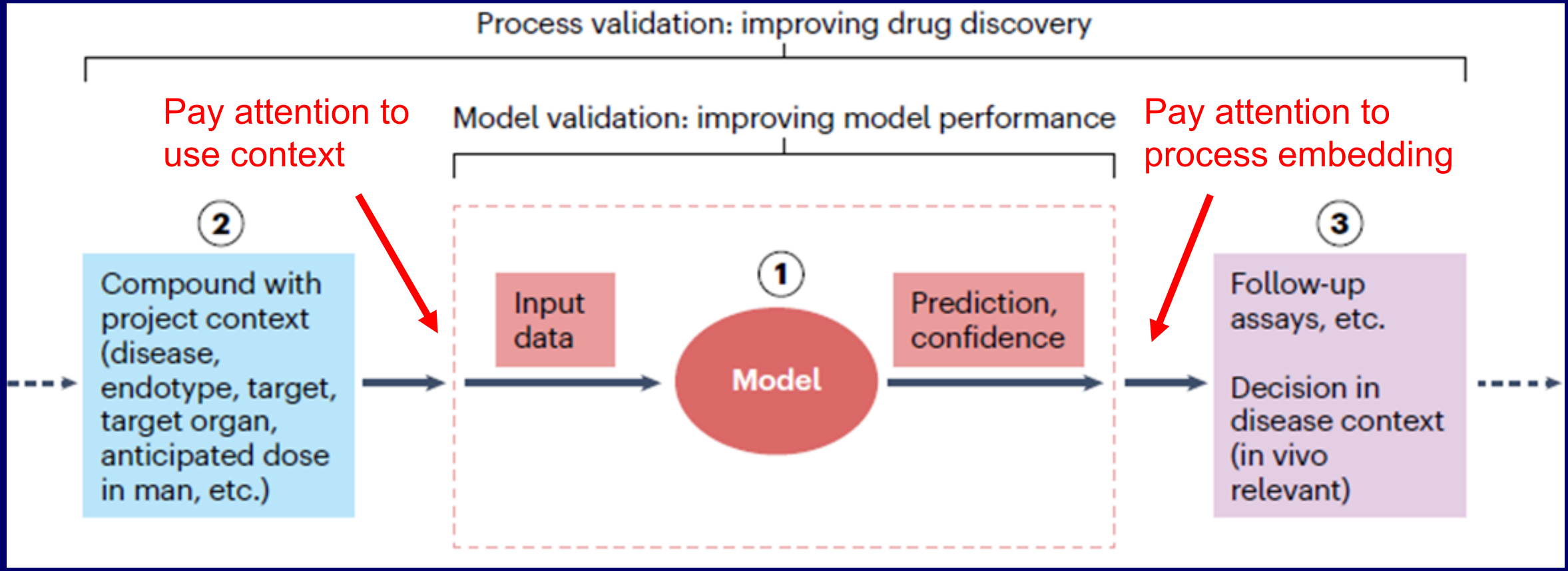
Type of Task	Primary Metric	Size of Space	High Score in Metric Achieved
Input	Recall	Medium	Can be high
Process	Accuracy	Low(-ish)	High (if explicit)
Decision	Precision	High	E.g. HTS: Can be high (high label predictivity, good heuristics)
		High	*In vivo*: Low (Low label predictivity, few heuristics)



Optimization and Execution (in predictive, low-dimensional spaces)

Judgement and Strategy (in high-dimensional spaces with low label predictivity)

4. Getting real: We don’t need ‘label predictors’, we need to impact the real world



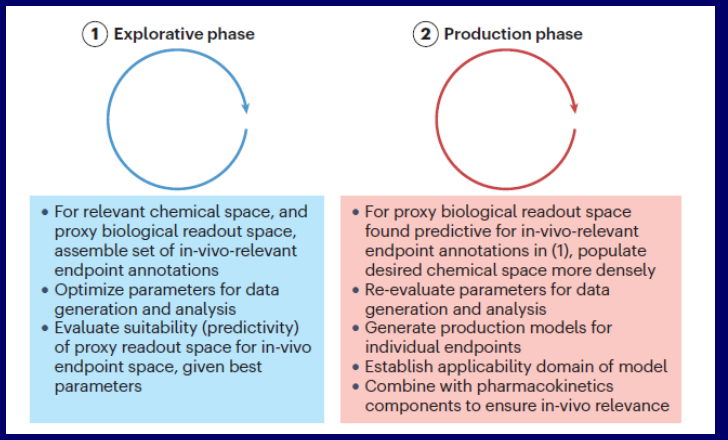
(a) S. Kearnes, Pursuing a Prospective Perspective, *Trends in Chemistry* 2021 (b) D’Amour, A. et al. Underspecification Presents Challenges for Credibility in Modern Machine Learning. *J. Mach. Learn. Res.* 23, 1-61 (2022) (c) Bender *et al.* Artificial intelligence in drug discovery — what it is, where we stand and the path forward. *Nature Reviews Drug Discovery* 2026 (25) 855–876.

Spaces are too large to be filled with data!

- Also (and in particular!) ‘spatially and time-resolved multi-modal multi-omics single cell (etc)’ data

- Hypothesis spaces can be large, but biases and heuristics help:
  - 10^{60} small molecules *but helpful heuristics and biases*
  - $>10^{9}$ target-disease-pharmacology-dose combinations (*in vivo*, not even considering more complex PK and safety) with *fewer helpful heuristics and biases*

Sweet spot: *Relevant biological systems + signal with precision + 'AI'*



1. Relevant biological context – say, human-derived co-culture/iPSC systems (needed because often we cannot model biology well enough)
2. Measurable from *relevant system* with *signal with sufficient precision*, *with relevance for decision (i.e. usually not from hypothesis-free screens!)*

AND THEN use

3. 'AI' for analysis

Explainer: 1. and 2. create helpful biases for 3. – lower-dimensional spaces, and better predictive value!

Bender *et al.* Artificial intelligence in drug discovery — what it is, where we stand and the path forward. *Nature Reviews Drug Discovery* 2026 (25) 855–876.

Necessary requirements for test system, data, model and validation to be practically useful

Requirements	Test system	Data	Model	Validation
For individual data points	Predictive for endpoint of interest	Sampling input space that covers space of interest; ground truth	Given test system and data -> needs to *generalize in feature space to ground truth*	Use case relevant ‘splits’ and metrics…
For distribution	Needs to provide learnable, generalizing input space for model	Needs to cover underlying ‘data generating process’		Plus ample prospective applications!

What did the old masters do right?
Sir James Black went *right to the matter of things*

James Black searches for a new type of histamine antagonist

Initially, three bioassay systems were selected by Jim. For in vitro assays, he chose the guinea-pig ileal muscle, which was the classical system for studying antihistamines such as mepyramine (now, of course, designated as an H_1 receptor system), and guinea-pig atrial tissue for the mepyramine-refractory histamine responses (where histamine increases the rate of beating of the right atrium). For gastric acid secretion there were no in vitro assays available and he selected the Ghosh and Schild procedure [5] (see below) in vivo. Subsequently, another in vitro assay was established using the rat uterus, in which contractions were evoked electrically, at one minute intervals, in a piece of rat uterine horn.

Personal reflections on Sir James Black (1924–2010) and histamine.
https://link.springer.com/article/10.1007/s00011-010-0269-2

1. Relevant system
2. Relevant readout
3. 'Data that delivers'

… while we now prefer to dance around the problem a bit ('tech', 'data', 'throughput', etc)

What creates value? Balancing economic benefit of decisions with ability to do so with data

1. Automated discovery where we have
   - *Low*-dimensional spaces
   - Data-***rich*** and smooth spaces
   - Where we can define (and evaluate) an ***objective function***

E.g. hit finding, optimization where SAR is available, materials discovery, …

2. Translationally relevant extrapolations in
   - **Already de-risked and hence lower-dimensional** decision spaces
   - Where data is available

Say, **good and fast** *fast*-followers

What matters most in commercial success: first-in-class or best-in-class?

Ulrik Schulze & Michael Ringel

Nature Reviews Drug Discovery **12**, 419–420 (2013) | Cite this article

Figure 1: Value captured as a function of time of market entry and therapeutic advantage.

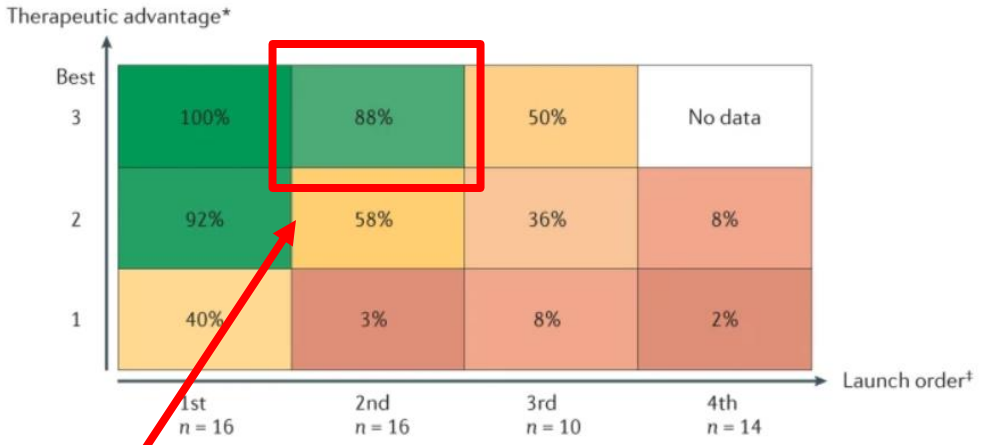
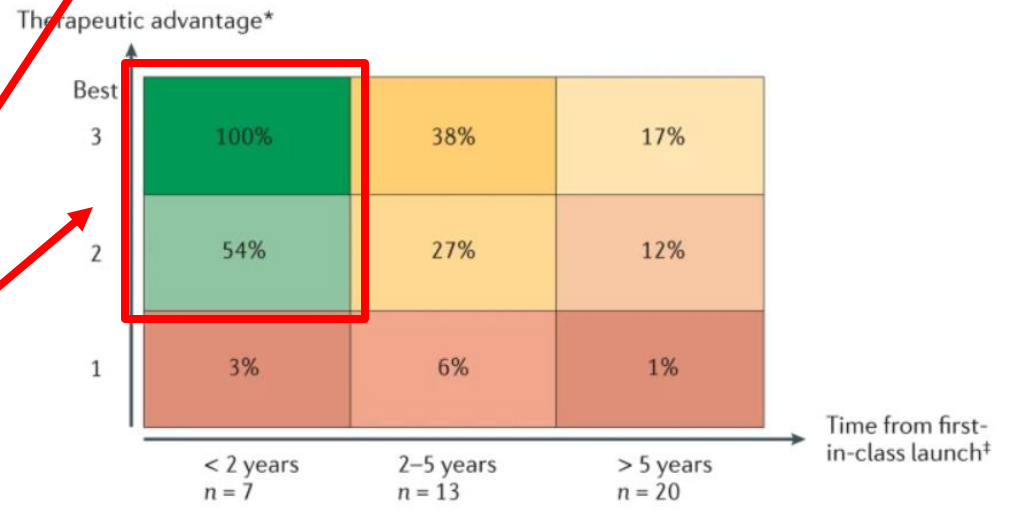


Figure 2: Value captured by follow-on entrants as a function of time of market entry and therapeutic advantage.



It is up to us what the future looks like

Data and understanding can lead us here!

Tech first and ‘data without a cause’ leads us here

Good	7 steps to solve a high-dimensional problem	Not so good
Sir James Black wants to find antihistamines	**1. Define problem in search and endpoint space**	‘We just generate data’, ‘we have a new tool’ etc
Sir James black starts with histamine	**2. Heuristics to make hypothesis space smaller**	Brute-force, high-throughput
Picks right assays (in vivo, predictive validity)	**3. Experimental setup with predictive signal**	‘Cheap and fast’
Makes analogues based on understanding *and* data	**4. Technically appropriate enumeration of search space**	Brute-force won’t win against numbers
(After trial and error) detects positive examples, w/ control	**5. Ability to detect positive signal**	Cheap doesn’t translate
(Assay is translational)	**6. Validate translational signal**	Failure in the clinic
Scale-up etc	**7. Productionize**	Project dead





Take-home message

1. Drug ‘discovery’ operates in *abundance setting*, *de-risking* (‘precision for clinical translation’) is what matters
2. We generally operate in data-poor environments (where ‘data poverty’ has many aspects; predictive value, dimensionality, *etc*)
3. Size of hypothesis space, predictivity of data for translationally relevant decisions, matter for *process* impact
4. Models help execution and optimization in spaces with predictive labels; less so for *data- and heuristics-poor decisions with low label predictivity*
5. Combination of data and knowledge; aligning understanding, data, model, ‘validation’, use case, and process integration determine success

Slides: www.drugdiscovery.net/broad26 - Contact: andreas@bio.bi