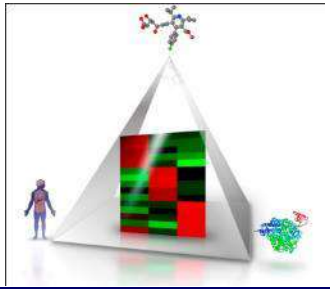
Artificial Intelligence in Drug Discovery: Where are we today? What else is needed to advance further?

Andreas Bender, PhD
Natural Philosopher for Molecular Informatics
Department of Chemistry, University of Cambridge

Director of Digital Life Sciences
Nuvisan, Berlin

Co-Founder of Healx, Ltd. and PharmEnable, Ltd.

UNIVERSITY OF CAMBRIDGE

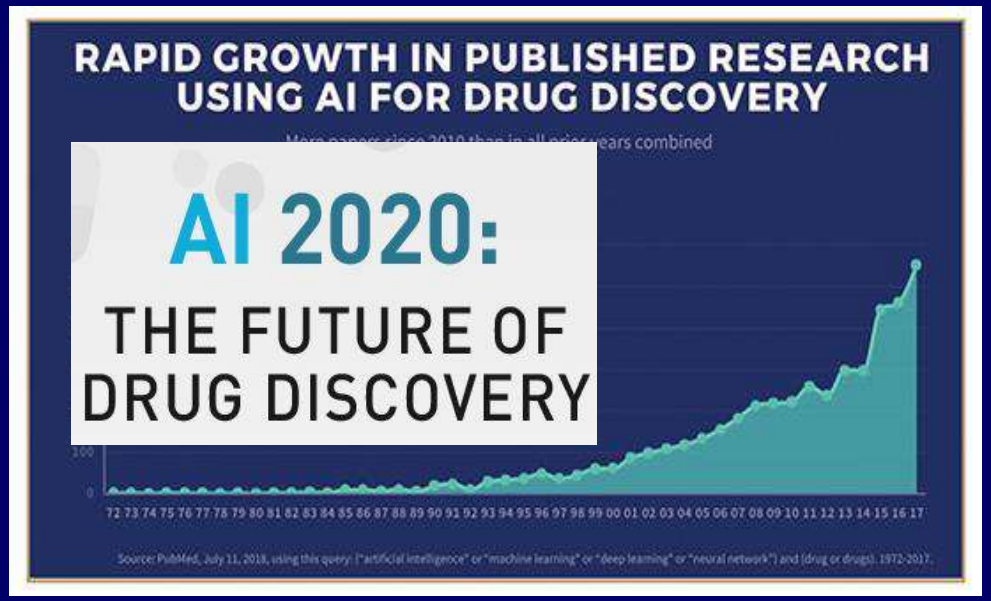


Any statements made during this talk are in my capacity as an academic

The 3rd wave of computers in drug discovery (80s, 2000, today) – time for realistic assessment has come

Fortune cover 1981

Recent headlines (2018-2020)



Old enough to remember 2000 biotech bubble, Human Genome Project, etc.

T. Reiss, Trends in Biotechnology, 2001:

“The number of drug targets will increase by at least one order of magnitude and target validation will become a high-throughput process.”

“More drug targets… 3,000–10,000 targets compared with 483”

Recent (2017) estimates of drug targets put the number currently at around *667*

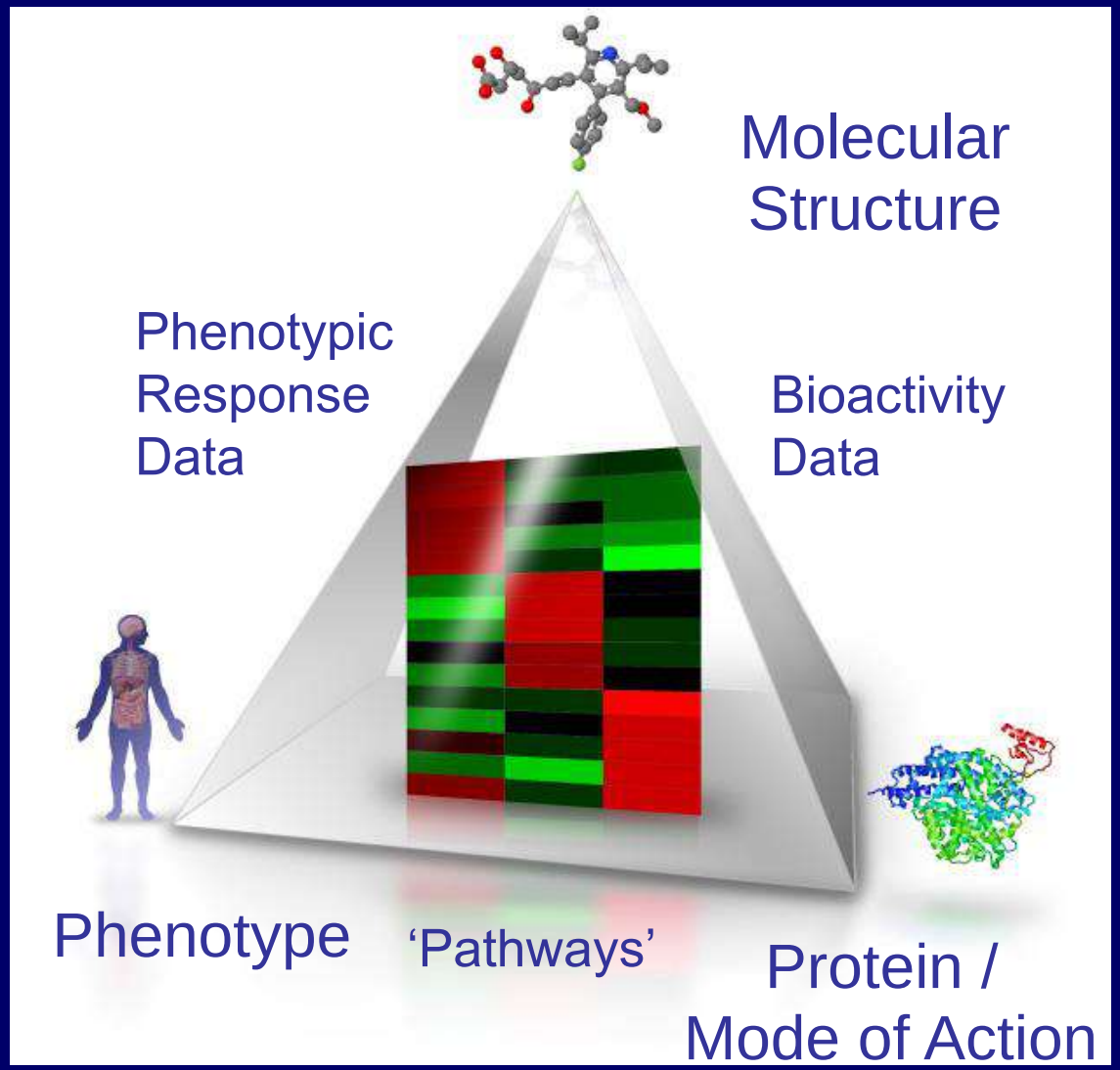
http://www.DrugDiscovery.NET/DataSignal

Outline: The data landscape, deep learning, biology… and humans

- Chemical and biological data: The flat-earth view
  - And where a flat earth is great!
- Chemical and biological data: The round-earth view
  - Drug discovery data and its complexity (… the elephant in the room…)

- Key learnings:
  1. The data we have is not the data we need
  2. … so what do we need, then?
  3. Model validation is poor….
  4. … and it is poor because of human biases, preferences

A simple view on the world: Linking Chemistry, Phenotype, Targets / Mode of Action (myself, until *ca.* 2010)



a.k.a.
"The world is flat"

= "We believe our labels"

(which are often insufficiently quantified, not directed or causal, unconditional, don't have time/concentration/ biological setup relevant for *in vivo* situation, *etc.*)

So what's the point of it all?
We would like to answer questions

- "What is the reason upon treatment with A for phenotypic effect B?"

  -> *Mode of Action*

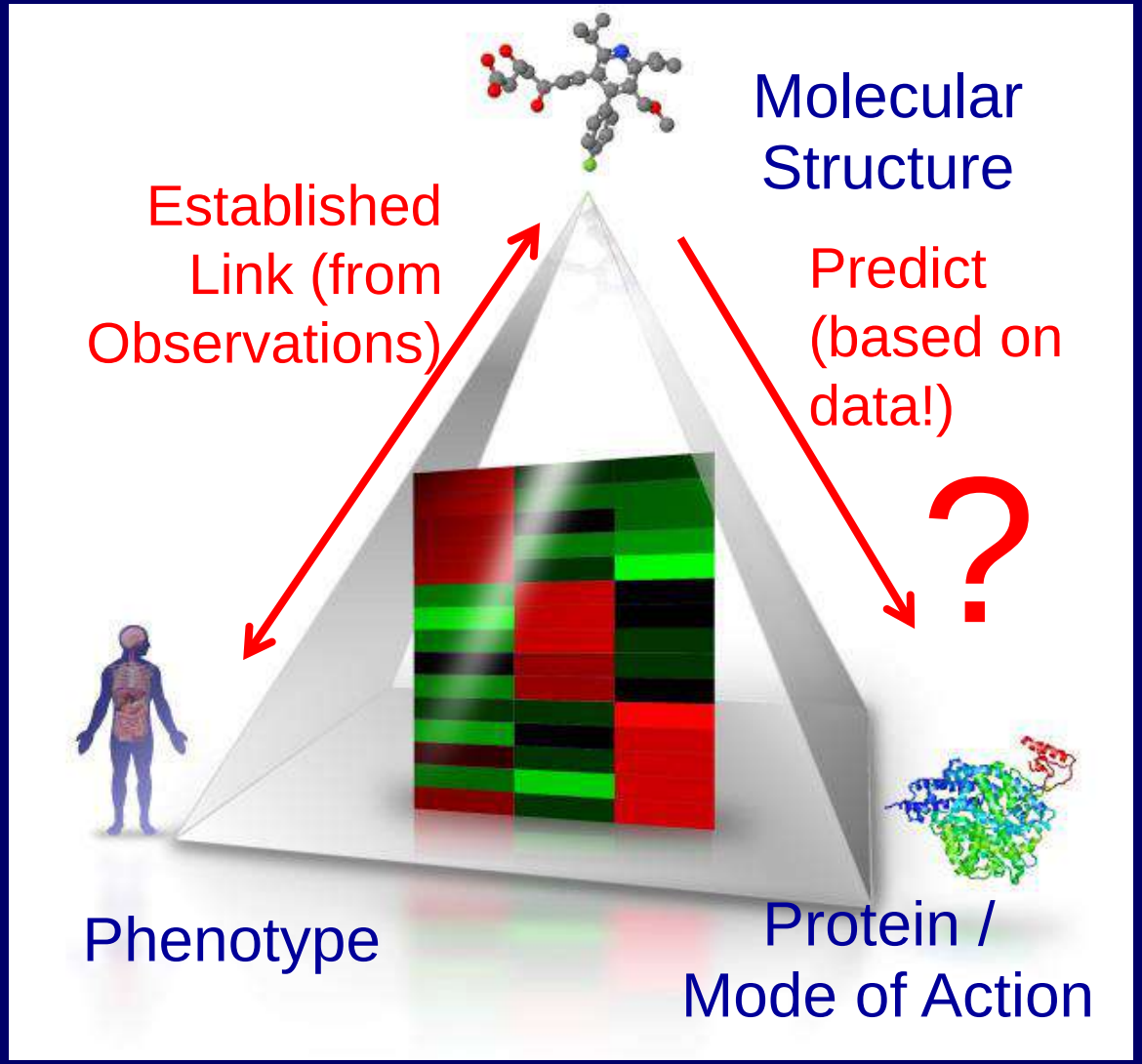
- "Which compound should I make to achieve effect C in a biological system?"

  -> *Chemistry*

- "Does patient D or patient E respond better to drug F?"

  -> *Phenotype / Phenotype Change*

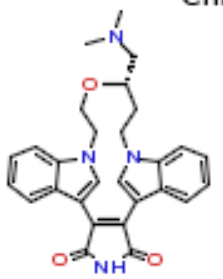
Starting from *in vivo* efficacy we can hypothesize the MoA, based on ligand chemistry

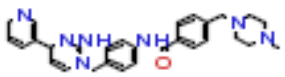


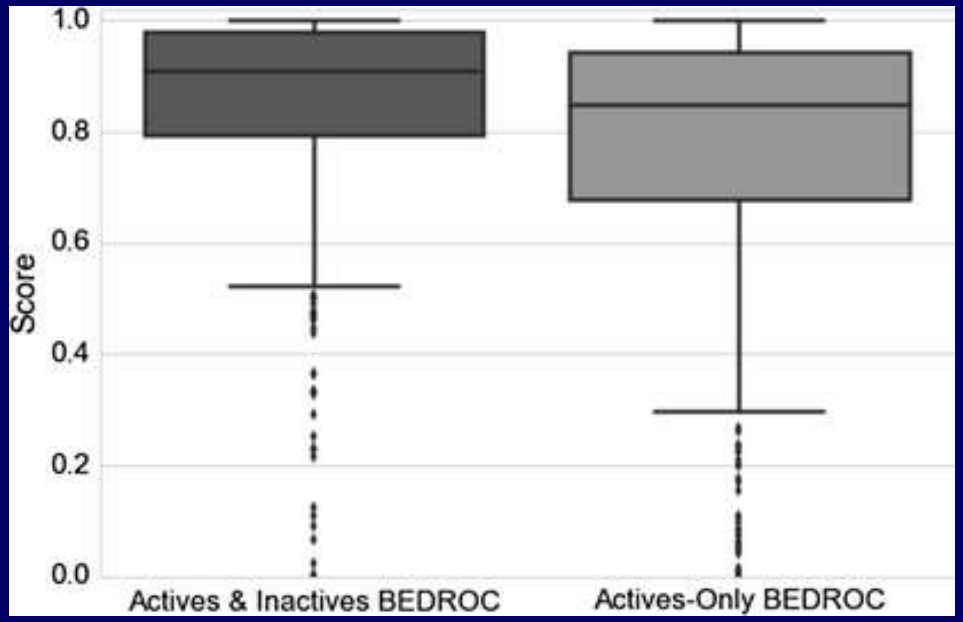
A. Koutsoukas *et* al., J Proteomics 2011 (74) 2554 – 2574.

The ‘flat earth’ view can *still* help! Eg Public target prediction model, based on ~200 mio data points

- E.g. work of Lewis Mervin, with AstraZeneca
- 2015, *J. Cheminformatics* (7) 51
- ChEMBL actives (~300k), PubChem inactives (~200m); 1,080 targets
- Can be retrained on in-house data
- https://github.com/lhm30/PIDGIN

Molecule	Targets	Scores
Chiral 	PRKCB1	95.81
	CAMK2G	87.48
	PRKCG	66.35
	PRKCA	56.99
	PRKCD	52.44
	PRKCH	51.41
	PRKCE	50.42
	PRKCZ	42.48

Molecule	Targets	Scores
	ABL1	46.50
	PDGFRB	28.99
	KIT	22.02
	CDK9	21.30
	BRAF	16.13
	FLT1	13.09
	PLK1	8.05
	BTK	5.44



Also data publicly available

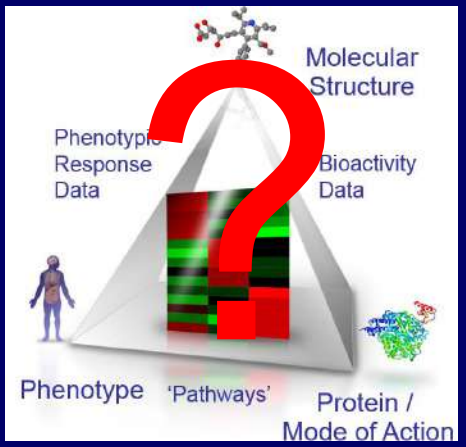
So: Using bioactivity data for ligand-protein activity modelling '*is relatively possible*'

- We make use of existing data (millions of data points!)
- On-target bioactivities (links between chemical structure and protein targets) are *relatively large-scale*, and *relatively homogenous*
- Hence, generating models for on-target bioactivities is 'possible'
- Can also be used for design (eg multi-target ligands)

BUT:

- Only covers known chemical space
- Suffers from various data biases (analogues, data set sizes, etc.)
- Labels are still heterogenous
- *In vivo* relevance of predictions needs to be established (!!!; PK, target engagement *in vivo*, competing ligand/knock-out, etc.)

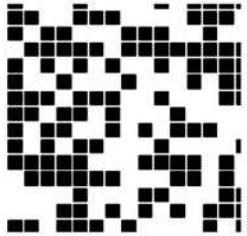
BUT…The world is not flat. What now?

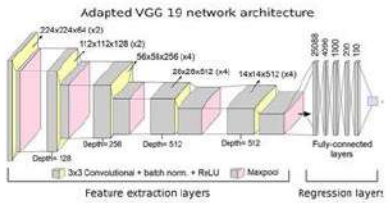


- Links between drugs/targets/diseases are quantitative, incompletely characterized
- Subtle differences in eg compound effects (partial vs full agonists, off-targets, residence times, biased signalling, etc.)
- 'Pathways' from very heterogenous underlying information; dynamic elements not captured etc.
- Effects are state-dependent (variation between individuals, age, sex, co-medication…) – PK is often rather neglected in AI approaches
- Phenotyping is sparse, subjective (deep phenotyping?)
- We don't understand biology ('the system'), we don't know what we *should* label, and measure, hence …
- We label what we *can* measure: 'T*echnology push*' vs '*science pull*' (!)
- **Are our labels – 'drug treats disease X', 'ligand is active against target Y', … - meaningful?**
- **Conditionality: Causality, confidence, quantification, ….?**
- **Computer science is tremendously powerful… but is our data?**

Object | Representation | Model | Object Label

Image Domain





ResNet?
AlexNet?
CapsuleNet?

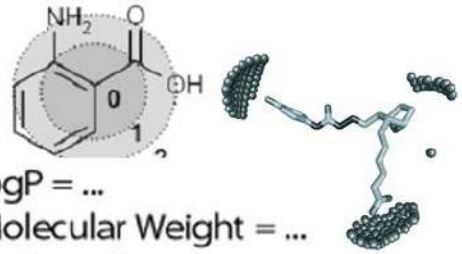
→ Cat

Representation and model are *intrinsically linked* (ie, model uses native object representation by pixels)

Largely *Unconditional* labels

Drug Discovery: Chemical Domain

?



logP = ...
Molecular Weight = ...
Molar Refractivity = ...

?

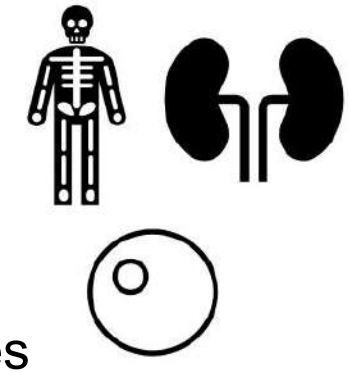
Artificial Neural Network/DNN? Support Vector Machine? Random Forest? Bayesian Classifier?...

→ Property A

Conditional labels (eg dependent on assay system, genotype, dose, endotype, sex, age, comedications, lifestyle, ...) ?

Both representation and modelling approach are *largely trial and error* (and *not* intrinsic to the chemical domain)

Drug Discovery: Biological Domain



?

Transcript-/proteomics? High-content imaging? Epigenetics? Histopathology? ....

?

Artificial Neural Network? Support Vector Machine? Random Forest?

→ State/Effect B

Heavily conditional labels (eg dependent on genotype, dose, endotype, sex, age, comedications, lifestyle, ...) ?

Both representation and modelling approach are *largely trial and error* (in particular the information content of biological readouts has only been established for particular cases)

Bender & Cortes
Drug Discovery Today 2021

Example of conditional labels: adverse reactions

- **“*Does* drug Y cause adverse reaction Z? Yes, or no?”**
- Pharmacovigilance Department: Yes, *if* we have…
  - A patient with this *genotype* (which is generally unknown)
  - Who has this *disease endotype* (which is often insufficiently defined)
  - Who takes *dose X* of *drug Y* (but sometimes also forgets to take it)
  - With known targets 1...n, but also unknown targets (n+1…z)
  - Then we see *adverse reaction (effect) Z* …
  - But only in x*% of all cases* and
  - With *different severity* and
  - *Mostly if co-administered with a drug from class C*, and then
  - More frequently in *males* and
  - Only *long-term*
  - (Etc.)
- **So – does drug Y cause adverse event Z?**

Data/'AI' in early discovery vs efficacy/safety

Early discovery/proxy space (usually *in vitro*)

- Often 'simple' readouts (eg protein activity), hence…
- Large number of data points for training models

- *Models have clear labels* (within limits of model system, eg 'ligand is active against protein at IC50<10uM', or solubilities, logP, or the like)
- Good for model generation: *Many*, *clearly categorized* data points

Efficacy/safety (usually *in vivo*)

- Quantitative data (dose, exposure, …)
- More complex models (to generate data), *fuzzy labels* (classes 'depend', on exposure, multiple eg histopathological endpoints) – hence…

- *Less, and less clearly labelled data*: Difficult from machine learning angle
- Data: *Recording* vs data *suitable for mining* – eg animal data tricky, even within single company

Problem setting in early discovery vs safety

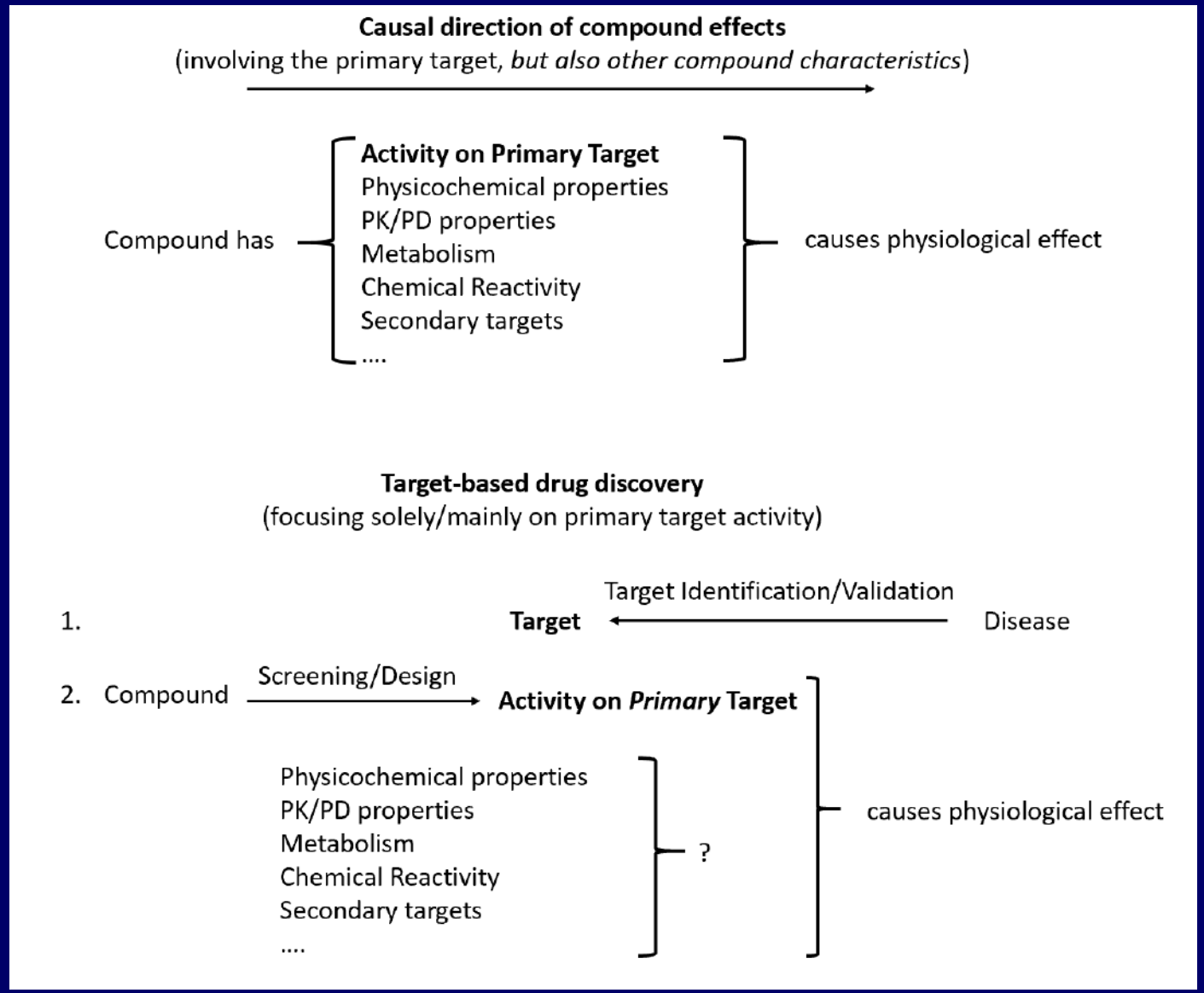
Early discovery/proxy space

- Discovery setting – ‘find me suitable 100s or 1000s out of a million’ (eg screening)
- Anything fulfilling (limited) set of criteria will do ‘for now’, predicting *presence of something*
- Computationally *generative* models often fine

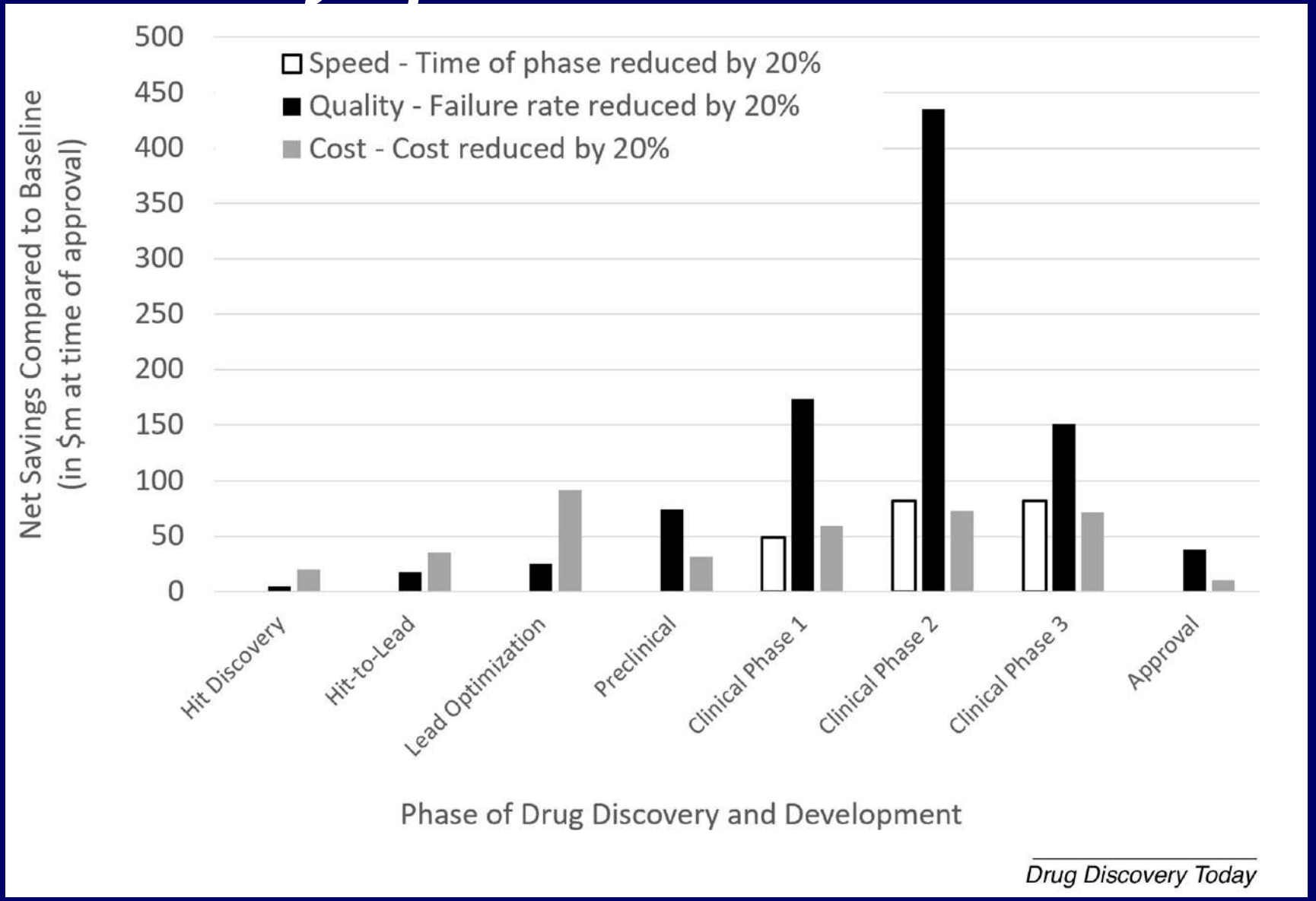
Efficacy/safety

- Need to predict for *this particular data point, quantitatively!*
- *Long list of criteria to rule out*, *based on limited data*… predicting *absence of ‘everything’* (eg different modes of toxicity)
- *Predictive* models (more tricky than generative!)

AI in drug discovery: Data availability drives the field of 'AI in drug discovery' … but a ligand is not a drug!



The *quality* of *in vivo-relevant* decisions matters more than *early speed*!



Discussion

1. The data we have is not the data we need
2. … so what data do we need, then?

3. Model validation is poor….
4. … and it is poor because of human bias

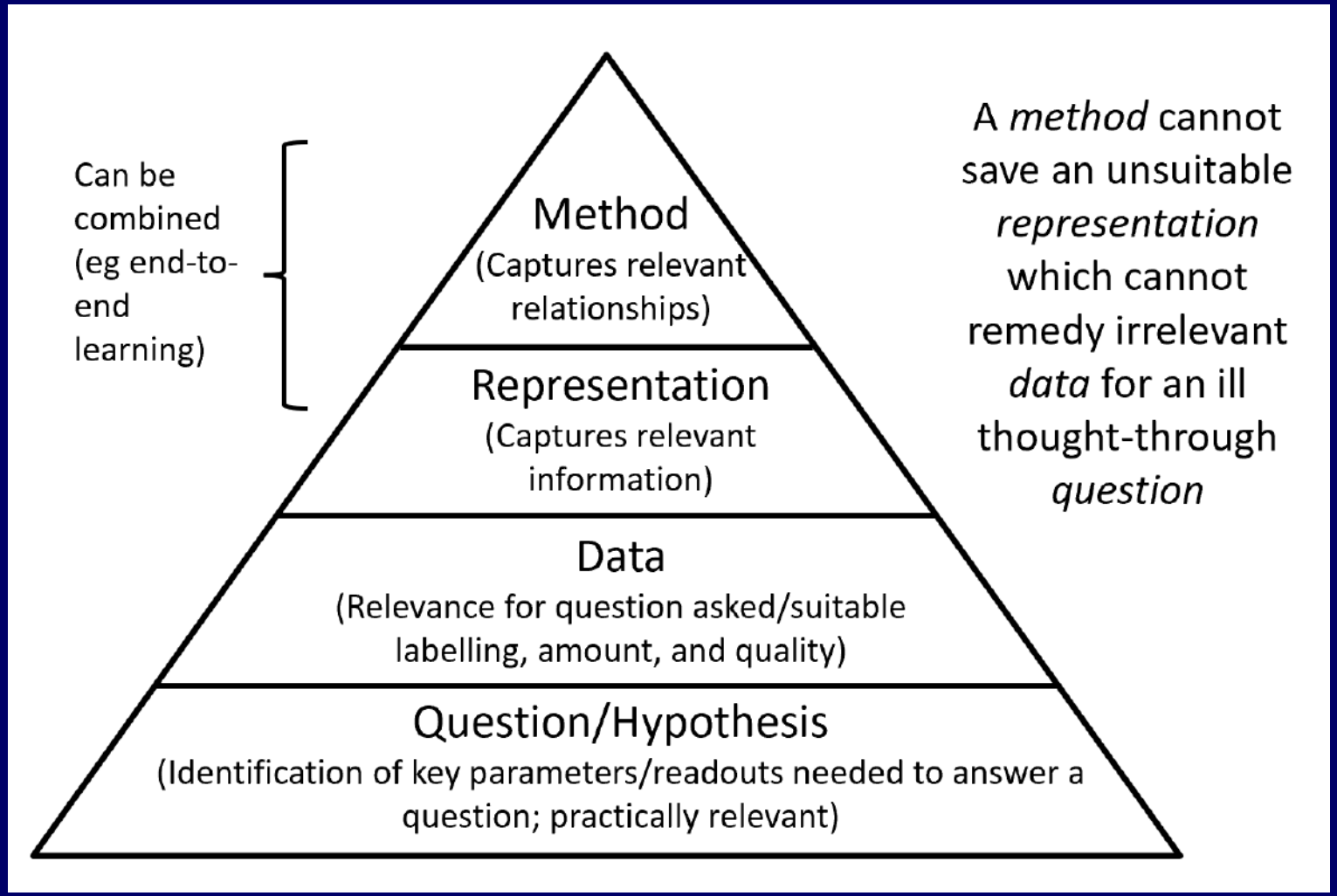
Much of the data we generate is generated for the wrong reasons (or in wrong ways)

- Often proxy measures (to reduce cost); historical data gets repurposed now ‘for AI’
- Not always relevant system/dose/time point/endpoint etc.

- “Models of models” – “the *in silico* model of the Glu/Gal mitotoxicity model” … is then meant to predict the *in vivo* situation
- We need to care more about modelling the actual endpoint of interest (say, organ risk), not the proxy (say, assay) endpoint!

- Often hypothesis-free (‘here we have our pile of data … anyone wants to have a go at it?’) instead of hypothesis-driven
- Often ‘technology push’, instead of ‘science pull’

The *question* needs to come first… and then the data, then the representation, and then the method
http://www.DrugDiscovery.NET/HowToLie



Lots of attention currently here…

But we need to care more about this

Validating a model *is not trivial*!

Model validation is always *process validation!*

- Training/test set split – too small, coverage irrelevant
- Prospective validation – too small, and biased (process!)
- Baseline model not well-chosen/optimized
- Data quality not assessed on context of model performance
- Relevance of model endpoint not assessed
- Result of *process* of validation ascribed to *model*

- ***"How to Lie With Computational Predictive Models in Drug Discovery"***
- ***http://www.DrugDiscovery.NET/HowToLie***

The bigger picture: ‘AI’ is where it is due in no small part due to human psychology

- Hype bring you money and fame – realism is boring
- FOMO (‘the others also do it!’) and ‘beliefs’ often drive decisions (‘maybe they *really* have the secret sauce?’)
- ‘Everyone needs a winner’ (‘*after investing X million we need to show success to the CEO/VP/our investors/…’)*
- Selective reporting of successes leads to everyone declaring victory (but in reality no one knows what’s actually going on)
- Difficult to really ‘advance a field’ with little real comparison of methods

What could make sense from the data side?

- We need *relevant* data (predictive for the *in vivo* situation), which is *possible to generate large-scale*

- ‘omics data: *Yes, but* experimental conditions (e.g. cell line)/dose/time point often don’t extrapolate to relevant situations
- Cellular morphology data: *Yes, but* we need to understand better what the applicability domain is/which interventions are visible in the readout
- Organ-on-a-chip: *Yes (!), but* still under heavy development, details to be seen

Summary

- We need to analyse our data (as we did for many years before), absolutely!
- ‘AI’/deep learning *is a valuable tool* in the toolbox
- The real game changer for translation to patients will come only once we understand biology/biological data better (and generate it, and encode it, and analyse it)
- Currently a lot of computer science-driven approaches, some of which are more applicable in drug discovery than others (real translation is necessary, *but also better experimental design!*)
- Consortia on even larger scale are needed (for targeted data generation, not just sharing what is there already)

Thank you for listening!
Any questions?

Contact: ab454@cam.ac.uk
Personal email: mail@andreasbender.de
Web: http://www.DrugDiscovery.NET
Twitter: @AndreasBenderUK