Damn the compass, full steam ahead: Decision theory and drug discovery

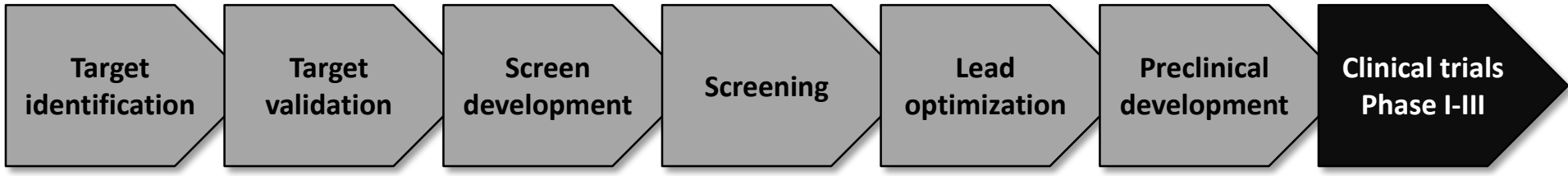
UKQSAR Spring 2017 Meeting

Jack.Scannell@ubs.com
Jack.Scannell@ed.ac.uk

Disclaimer

The views expressed in and during this presentation are those of Jack Scannell. They do not represent those of UBS.

How Experts Describe Drug R&D to Bankers Like Me



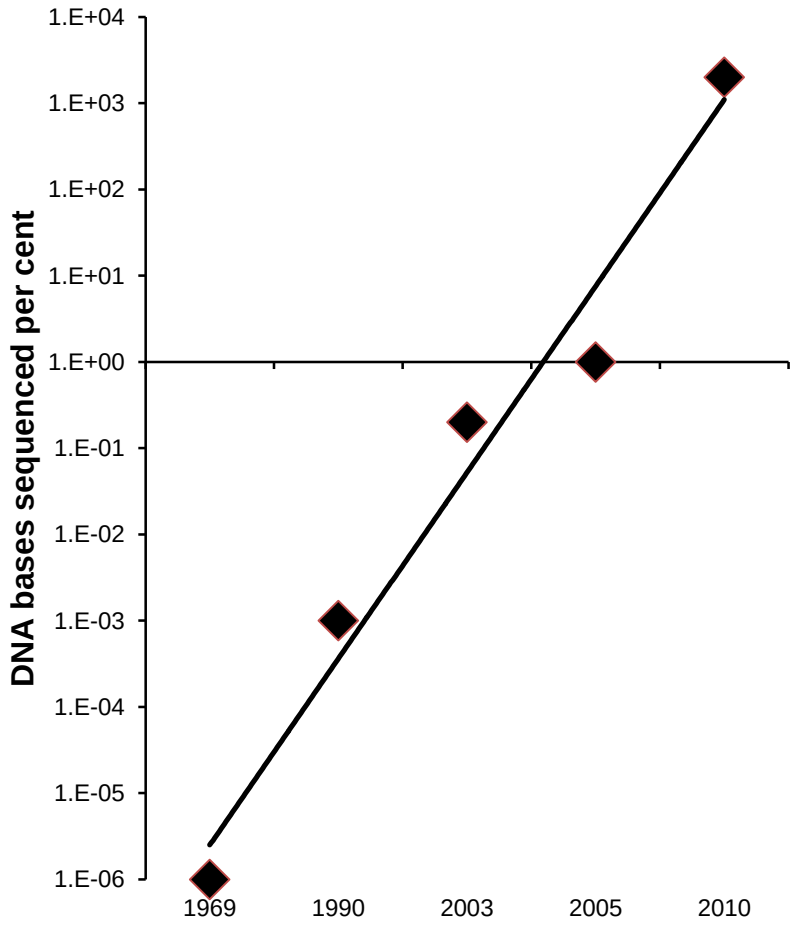
The aspiration that has been industrialized:

- Molecular component ***x*** (mis)behaves causing disease phenotype ***y***
- Molecular component ***x*** can be drugged with ***d*** in a way that causes an improvement in ***y*** without unacceptable side effects or toxicity
- There still exist many identifiable, exploitable, instances of ***x***, ***d***, & ***y***
- **Therefore** the industrial process set out above will deliver, at low cost, a large number of successful drug candidates into clinical trials

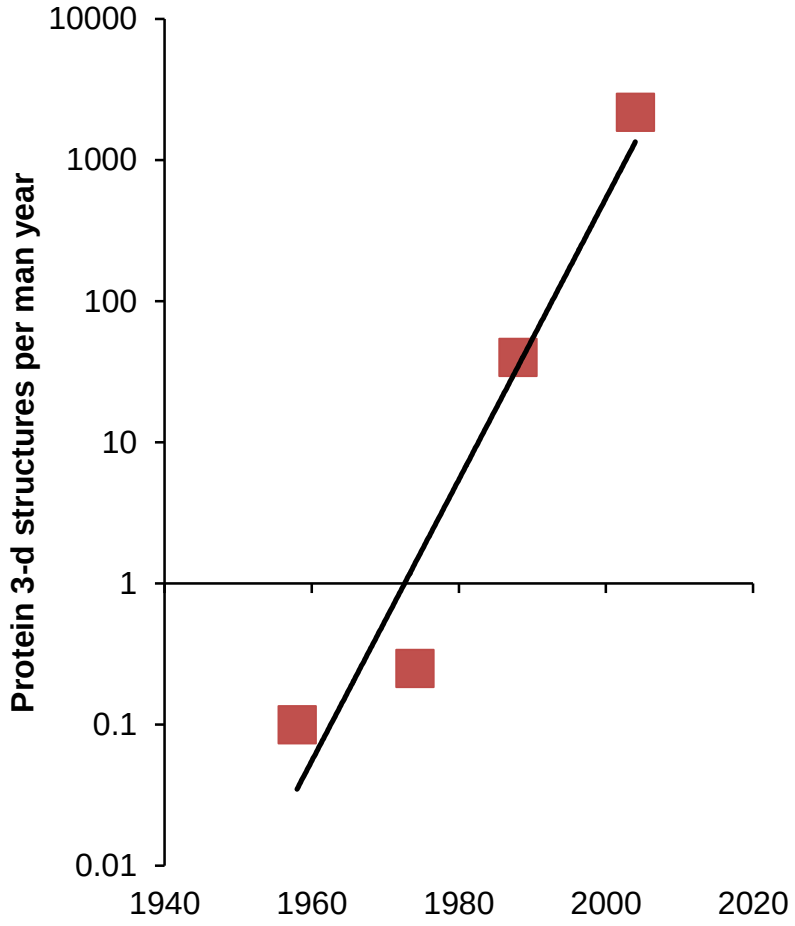
3

R&D Inputs Are Faster, Better and Cheaper than Ever

10^{10}x change in DNA sequencing efficiency?



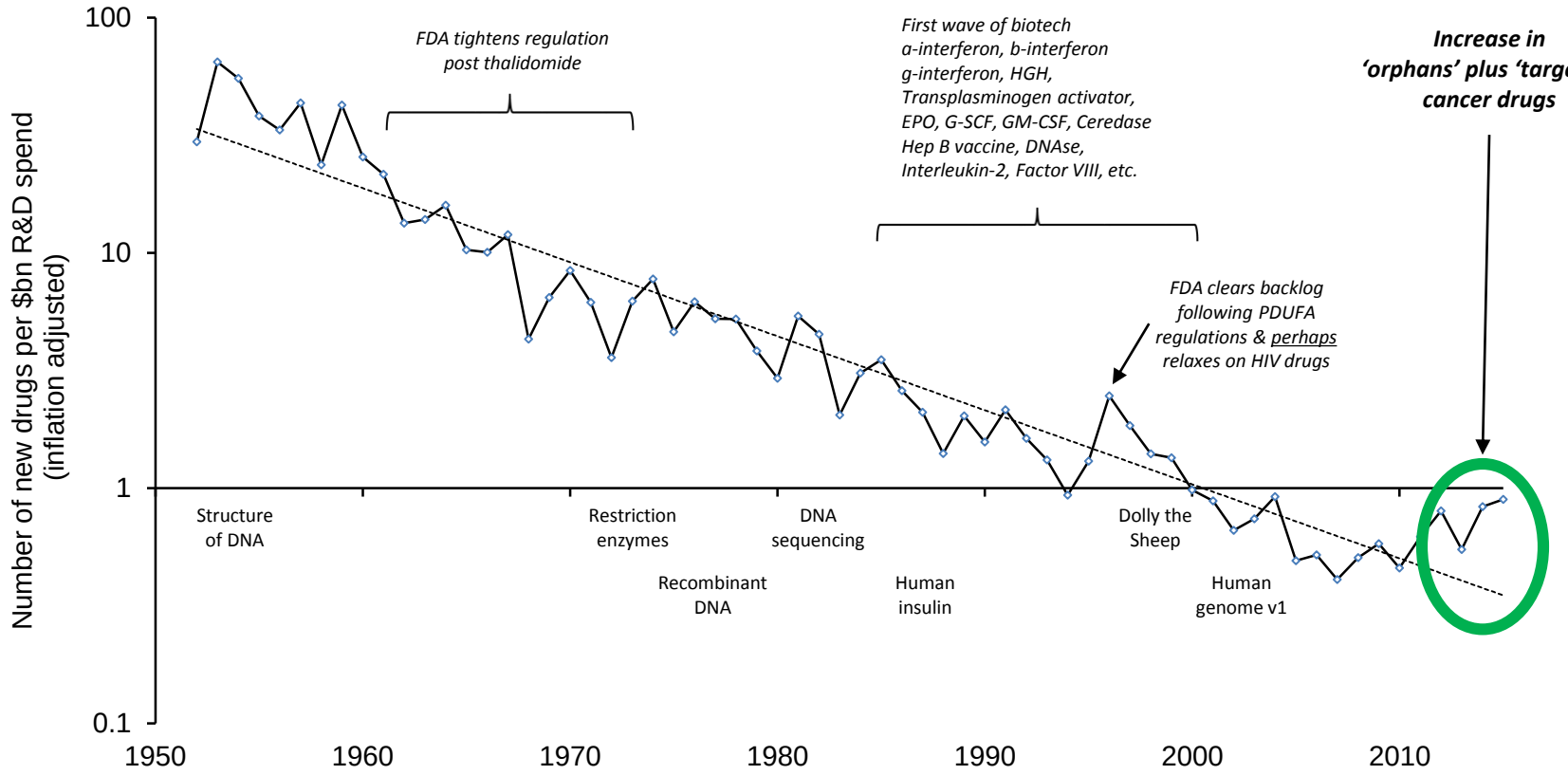
10^{4}x change in protein X-ray crystallography efficiency?



Source: Bernstein analysis, Hogan 1997, Geysent et al, 2003, Dolle 2011, Sanger 1988, Meldrum et al, 2011, Joachimiak 2009, Van Brunt 1986, Mayr & Fuerst 2008.

4

But The Number of New Drugs Approved Per $bn of R&D Spending Has Fallen by Around 100-fold Since The 1950s



Note: R&D costs are based on the PhRMA Annual Survey 2011 and reference Munos 2009. PhRMA is a trade association that does not include all drug and biotechnology companies so the PhRMA figure understates R&D spending at an industry level. Total industry expenditure since 2004 has been 30% to 40% higher than the PhRMA members' total spend, which formed the basis of this figure. The NME count, on the other hand, is the total number of small molecule and biologic approvals by the FDA from all sources, not just PhRMA members. The overall picture seems fairly robust to the precise details of cost and inflation calculations. Source: Munos 2009, PhRMA annual surveys, FDA, and Bernstein analysis.

Similar Efficiency Trends in the Whaling Industry circa 1850

Merchant shipping became faster, better, and cheaper

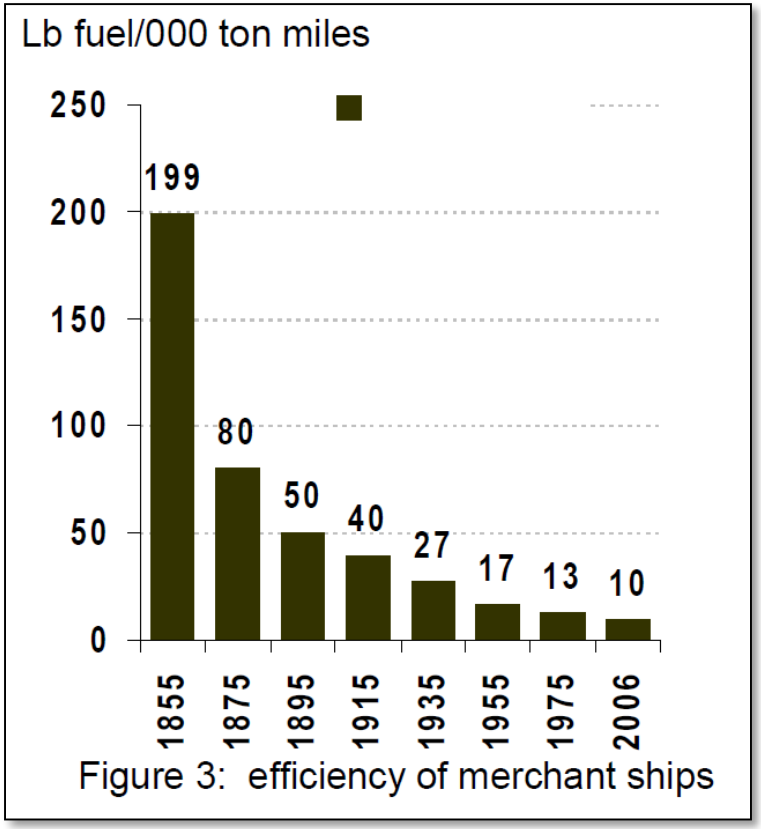
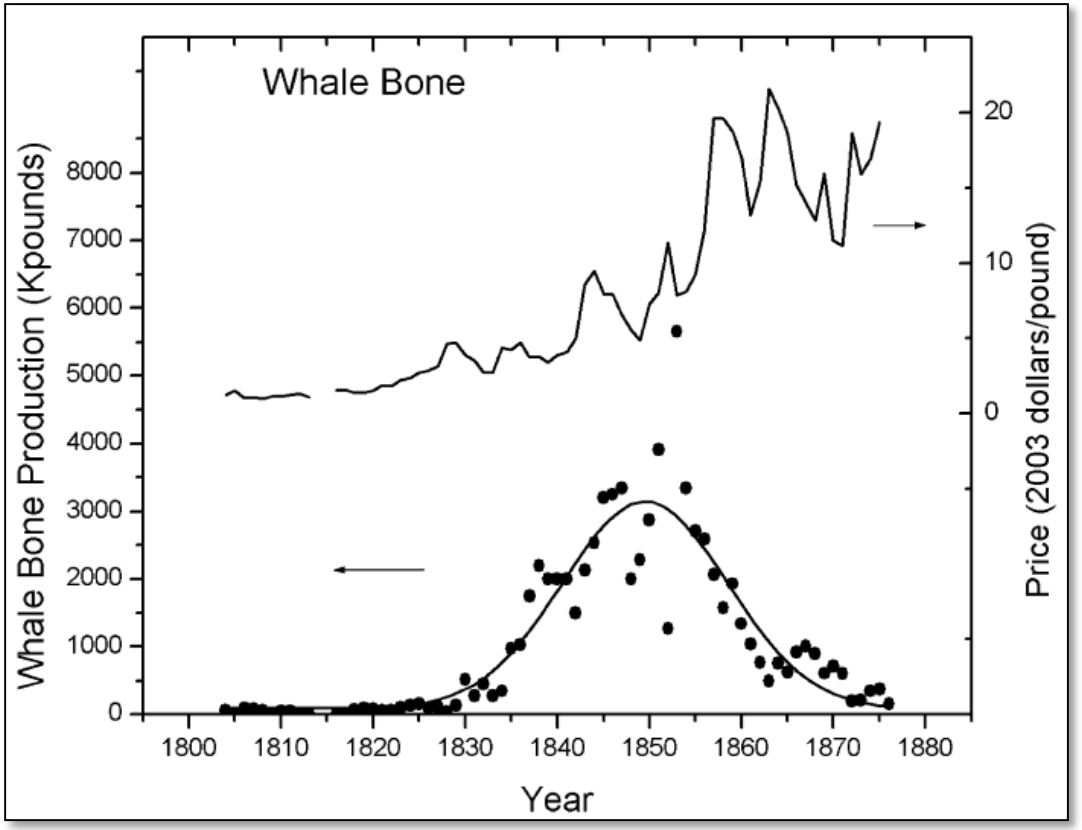


Figure 3: efficiency of merchant ships

Source: Stopford (2010). How shipping has changed the world, and the social impact of shipping

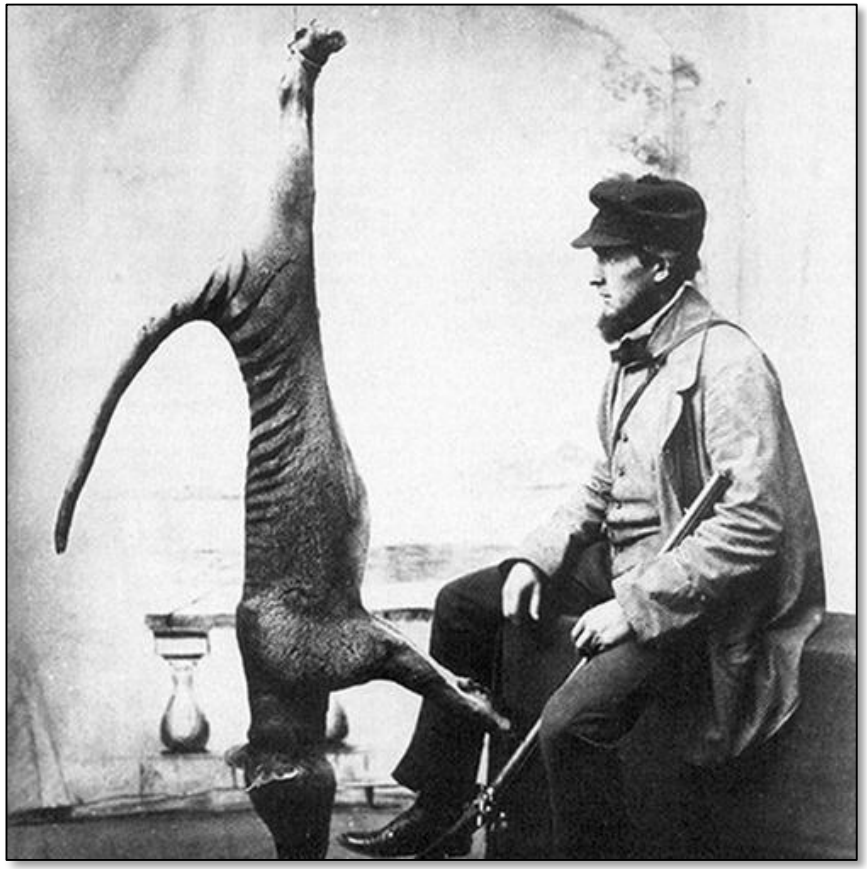
Whale bone became much more expensive to produce



Source: Bardi (2007). Energy prices and resource depletion: Lessons from the case of whaling in the nineteenth century

Two Possible, Non-Exclusive, Explanations For The Contrast Between R&D Input and Output Trends

Progressive depletion of tractable opportunities



Progressive adoption of less efficient R&D methods

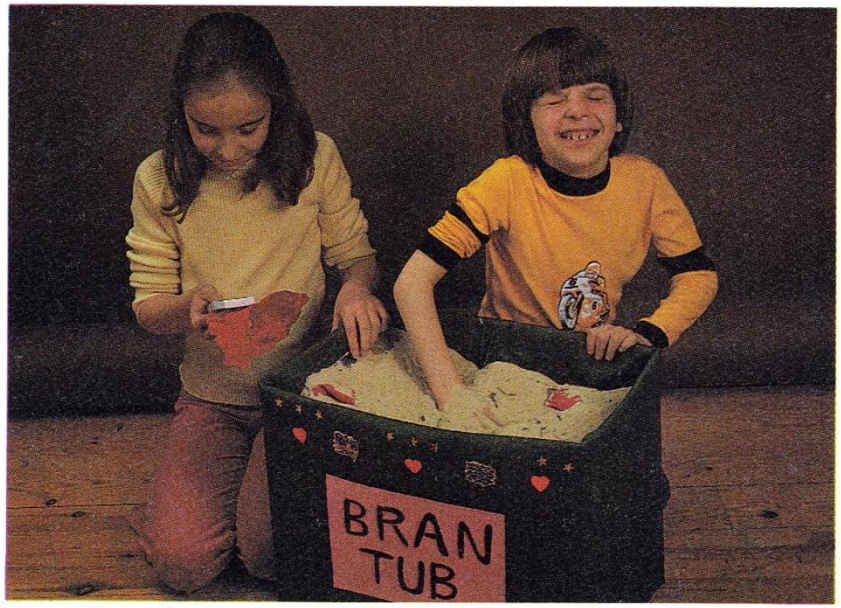
Source: Wiki Commons/public domain.

Different Kinds of Search Respond Differently to Brute-Force Efficiency

Children depleting a bran tub

Party games Paper crafts

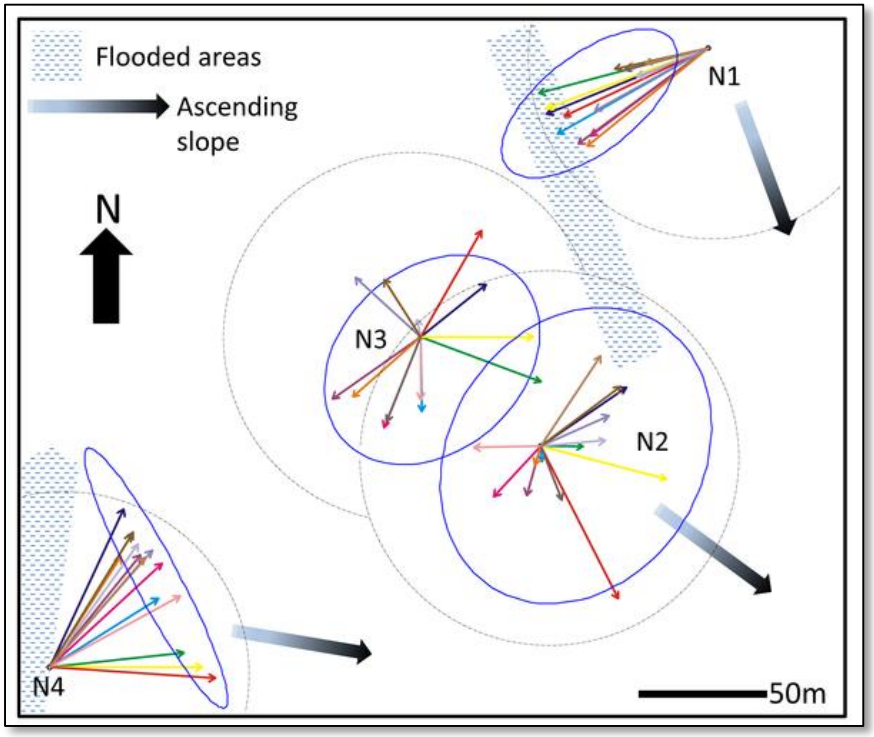
Bran tub **570**



You will need:

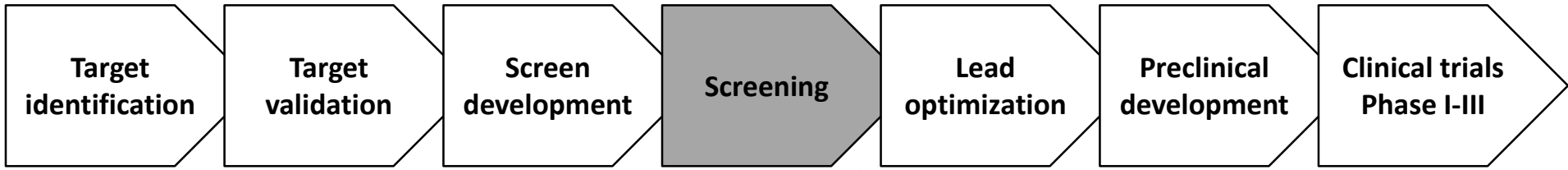
Large cardboard box or wooden or plastic container
Sawdust or sand
Presents
Crêpe paper
Wrapping paper (optional)
Paper for notices
Motifs for decoration
Scotch tape

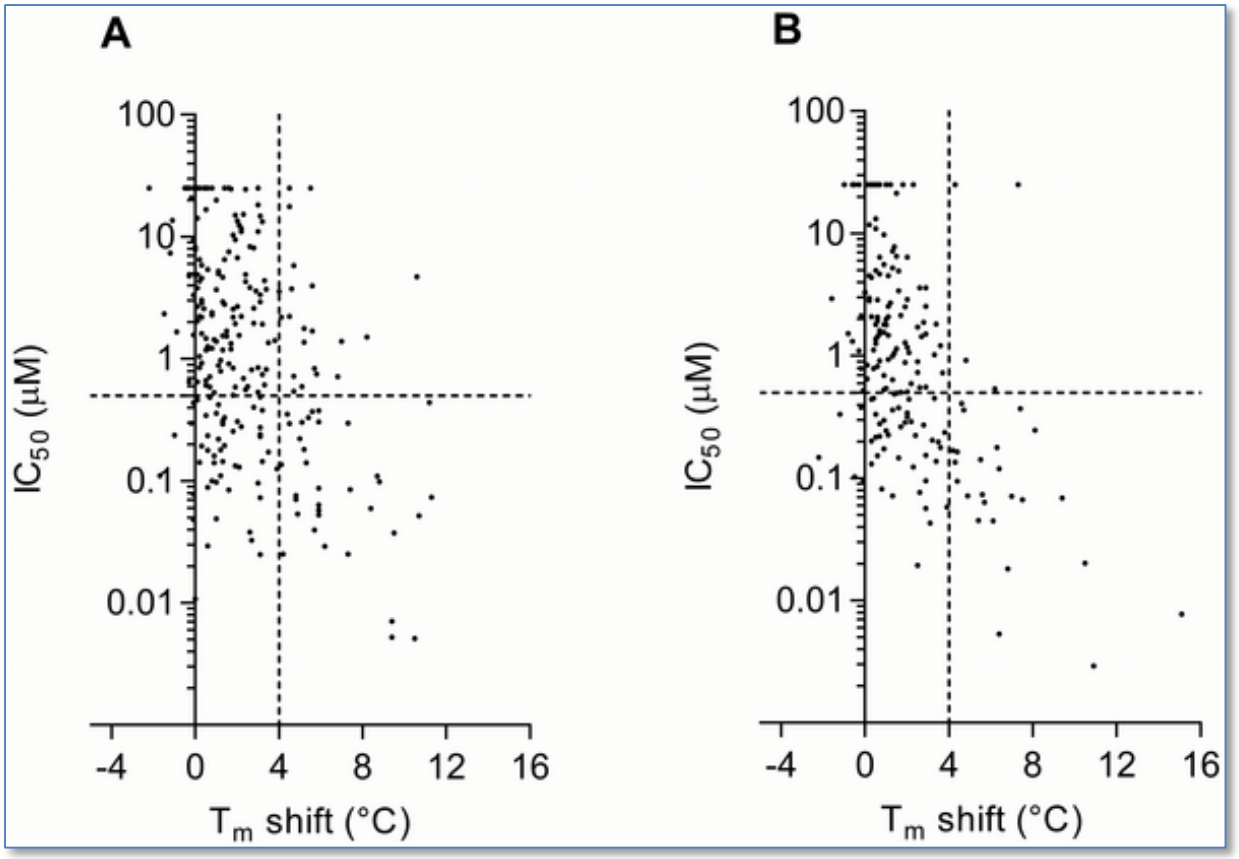
Ants depleting a grassland



Source: Lopes et al. (2016) Spatio-Temporal Dynamics of Foraging Networks in the Grass-Cutting Ant Atta bisphaerica Forel, 1908 (Formicidae, Attini). PLoS ONE 11(1): e0146613. doi:10.1371/journal.pone.0146613

Efficient Search Also Depends on Heading in the Right Direction

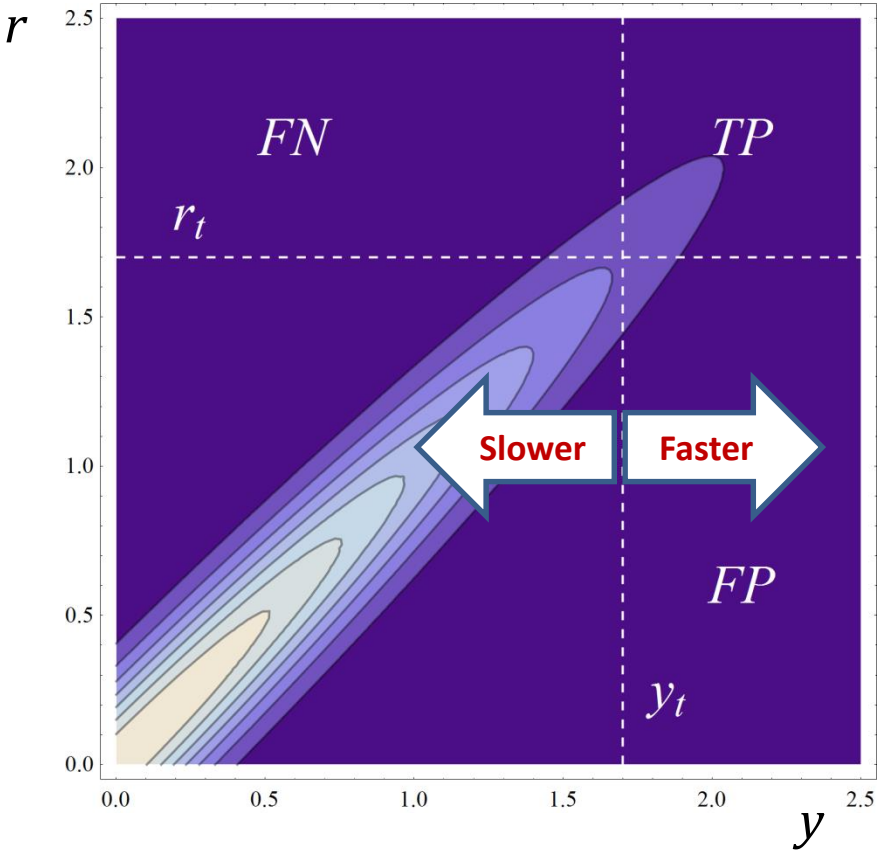




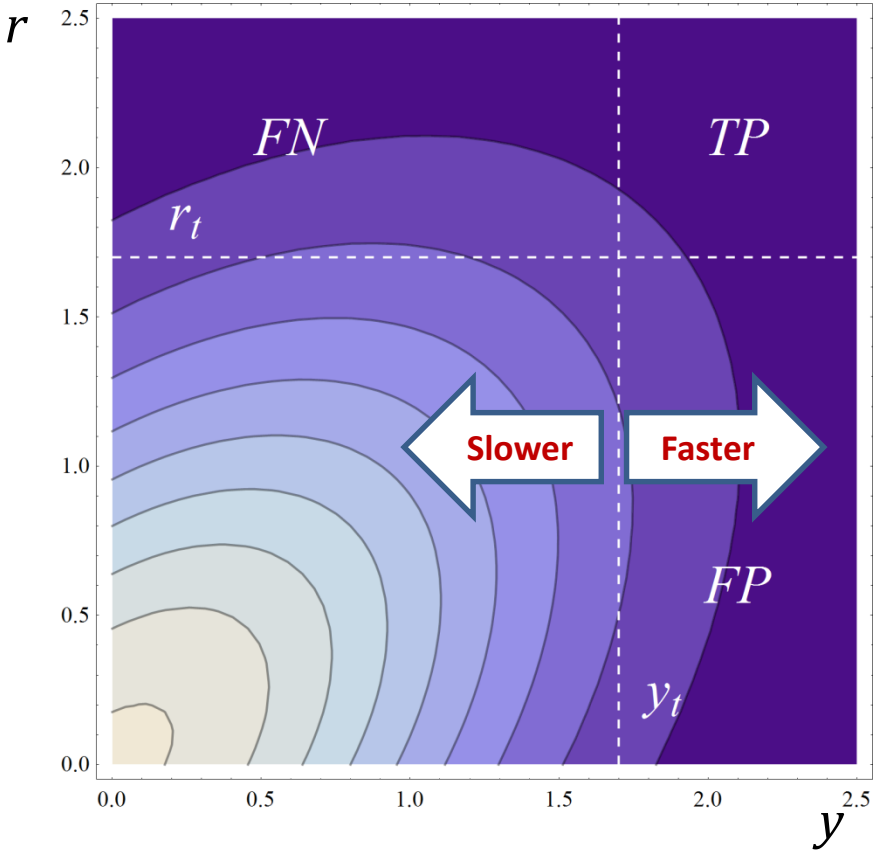
Rudolf et al. (2014) A Comparison of Protein Kinases Inhibitor Screening Methods Using Both Enzymatic Activity and Binding Affinity Determination. PLOS One. http://dx.doi.org/10.1371/journal.pone.0098800

A Decision-Theoretic View of a Screening or Disease Model

Good model (high predictive validity)



Bad model (low predictive validity)

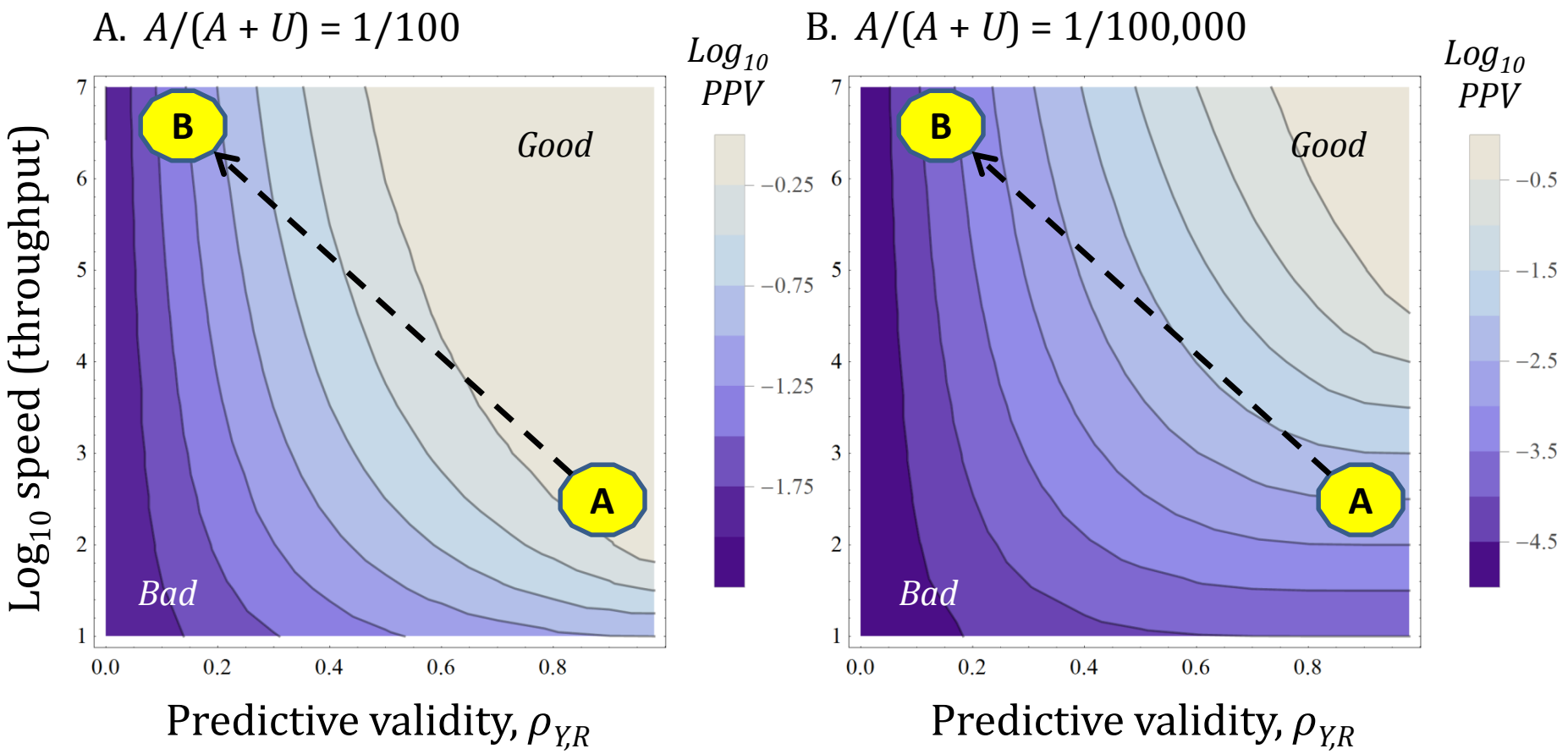


Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

*"Is it not **peculiar** that the first useful antibiotic, the sulphanilamide drug prontosil, was discovered by Gerhard Domagk in the 1930s from a small screen of available dyes (probably no more than several hundred), whereas screens of the current libraries, which include ~10^7 compounds overall, have produced nothing at all?"*

Lewis K. *Plaforms for antibiotic discovery.* **Nat. Rev Drug Discov.** 2013; 12: p. 371-387

It isn't Peculiar, if We Replaced Highly Valid Screening and Disease Models (A) With Much Less Valid Ones (B)



Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Counter-intuitive Navigation in High Dimensional Spaces

4096 Points in an n-dimensional grid with unit spacing between points…..

Number of grid dimensions	Grid layout	Median # neighbouring points	Corner to corner distance
1 (line)	4096	2	4096 units
2 (square)	64 x 64	8	90.5 units
3 (cube)	16 x 16 x 16	26	27.7 units
4 (hypercube)	8 x 8 x 8 x 8	80	16 units

The Blackian Demon Beats Brute-Force Search if You Can Step Through High Dimensional Chemical Space in the Right Direction

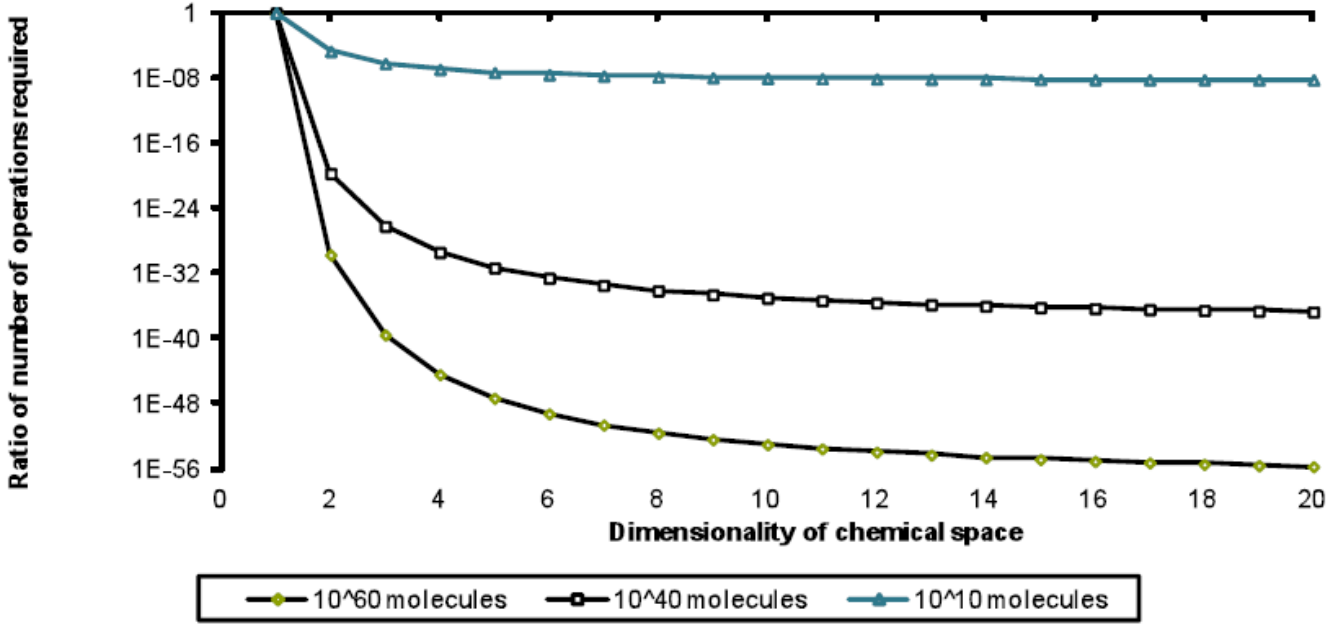


Figure | **Ratio of the number of steps required by the Blackian demon versus super-HTS to exhaustively search a space**. | The lines show the ratio of the number of steps required by the Blackian demon versus the number of steps required by super-HTS to exhaustively search for the best molecule in spaces of varying dimensionality (horizontal axis) in molecule universes of varying size (lines). Note the logarithmic vertical axis. If chemical space contains a very large number of molecules, and if the space is high dimensional, the Blackian demon requires many orders of magnitude fewer steps to find the best molecule.

Source: Scannell, Blanckley, Boldon, & Warrington 2012. Nature Reviews Drug Discovery 11, 191-200 (March 2012) | doi:10.1038/nrd3681. See supplementary materials

Why "Common Sense" Assumptions about R&D Productivity Could be Wrong

1. The “standard model” of drug R&D appeals to common sense via survivor-bias and post-hoc rationalisation
   - Imatinib/Gleevec is to the “standard model” what the human eye is to “creation science.”
2. But common sense is a bad guide to the quantitative importance of model validity when positives are rare
3. And ommon sense is a bad guide to the superiority of directed iterative search over brute-force search in high dimensional spaces
   - Provided models are valid enough to point in the right direction

Projects that Could Follow from This Work

Test the validity hypothesis and allocate the blame

1. Build an "historical atlas" of screening and disease models from 1950 to 2016
2. Identify 3 to 5 representative therapy areas
3. Track quantitative model performance over time
4. Assess the degree to which predictive validity (or proxies thereof) has declined
5. Allocate blame to exhaustion and retirement, versus changing fashion, etc.

Work with decision-makers to steer investment toward valid models

1. Show how project outcomes are sensitive to predictive validity
2. Develop a lingua franca for validity-related concepts for groups of R&D decision makers
3. Develop frameworks to evaluate the likely predictive validity of models in competing proposals
4. Learn how to ask for the right information from applicants
5. Integrate validity assessment with other constraints (e.g., project cost)

Thanks to the Following Individuals for Their Help With….

Academia

- Joyce Tait & Richard Barker

Statistics

- Brendan Jackson, Henry Stott, James Geddes, Mark Latham & Kostas Paraschakis

Collaboration

- Jim Bosley

Drug discovery

- William Bains, Nicholas Edwards, Brian Warrington, Geoff Lawton, Chas Bountra, Paul Brennan, & Stephan Knapp

Any Question?

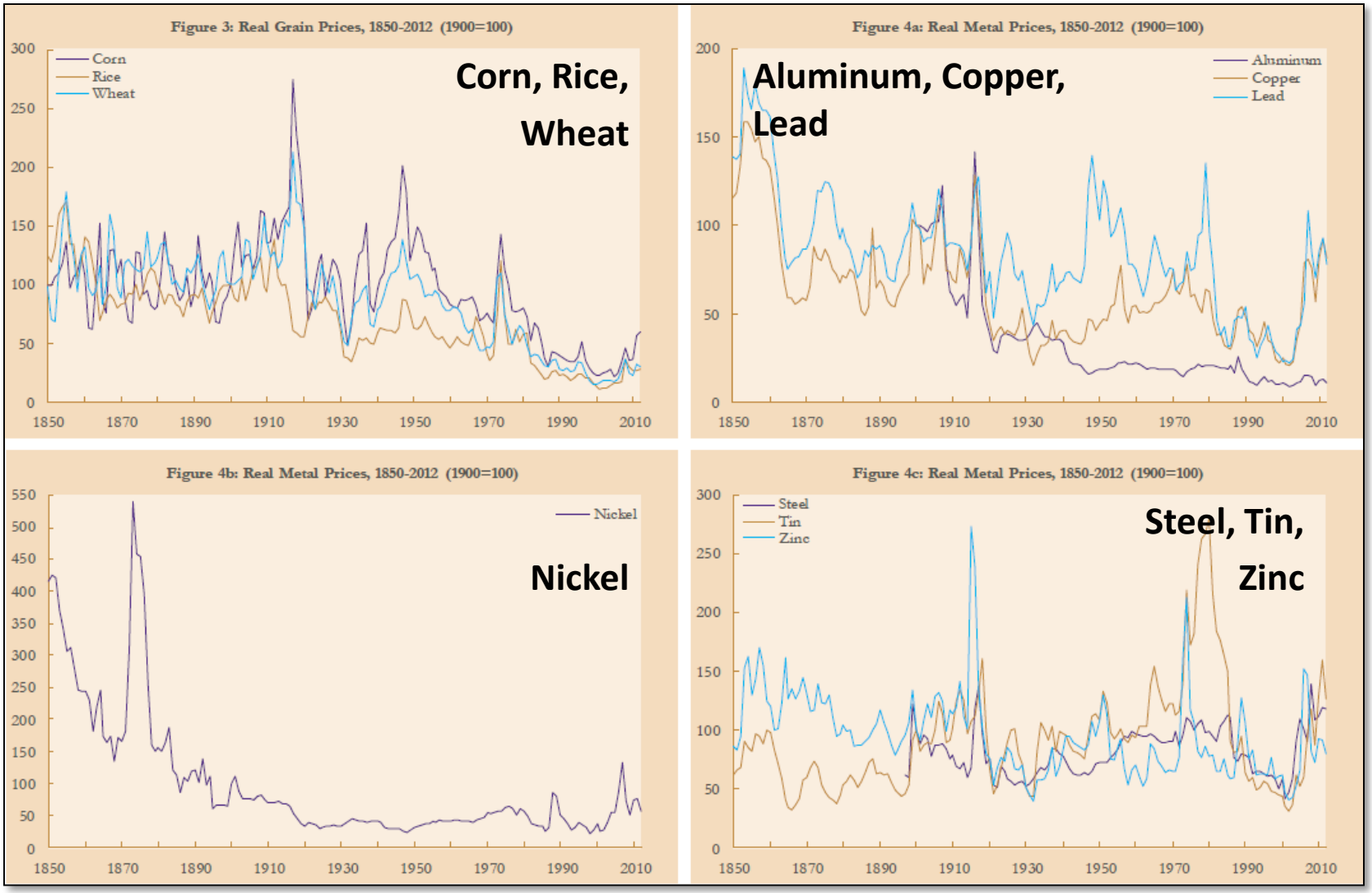
UNUSED SLIDES FOLLOW

Background Reading

- Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B. (2012). *Diagnosing the decline in pharmaceutical R&D efficiency.* Nature Reviews Drug Discovery, 11(3), 191–200. doi:10.1038/nrd3681
- Scannell, J. W., & Bosley, J (2016). *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215
- Scannell, J. W. (2015). Pharmaceutical evolution: *Clinical selection versus intelligent design.* ABPI UK Biopharma R&D Sourcebook 2015, reprinted as Innogen working paper no. 115

Production Costs Actually Decline in Many "Extraction" Industries

Commodity Prices 1850 to 2014



Source: Jacks (2013) From boom to bust: A typology of commodity prices in the long run.

Decision Terminology (I)

Decision variable, generally *Y* or *y*	Measures on which R&D progression decisions are based
Reference variable, *R* or *r*	Measures which test of the performance of the decision process
Decision threshold, generally y_t	The decision is "*yes*" when $y \geq y_t$ and "*no*" when $y < y_t$
Reference threshold, r_t	Something is a positive when $r \geq r_t$ and a negative when $r < r_t$
Predictive model, PM	Something that generates decision variables for therapeutic candidates
Predictive validity, PV	The degree to which the ordering of the population of candidates on the decision variable would match the ordering of the candidates on the reference variable, when sample sizes are large

Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

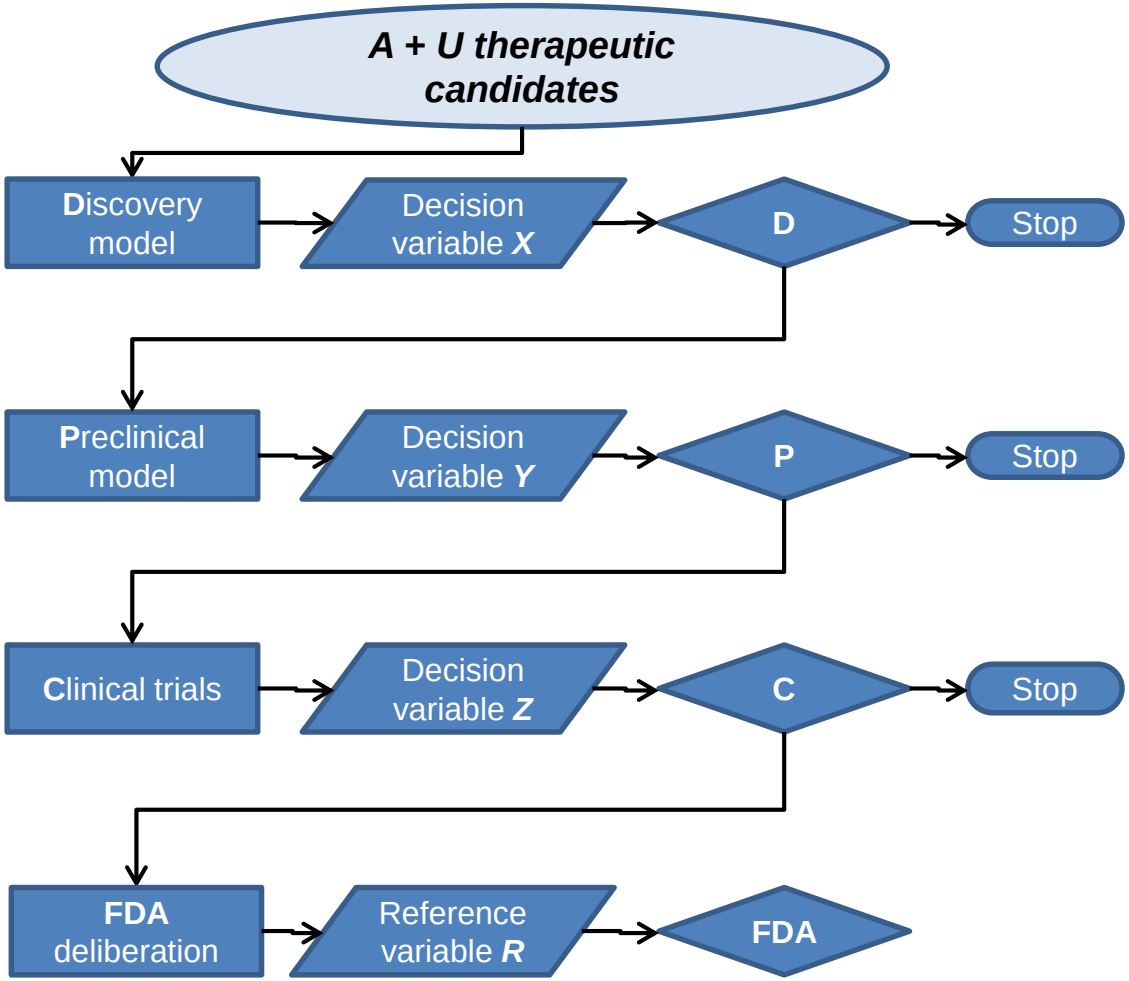
Decision Terminology (II)

True positives, *TP*	Items for which $y \geq y_t$ and $r \geq r_t$
True negatives, *TN*	Items for which $y < y_t$ and $r < r_t$
False positives, *FP*	Items for which $y \geq y_t$ and $r < r_t$
False negatives, *FN*	Items for which $y < y_t$ and $r_i \geq r_t$
True positive rate, *TPR*	$TPR = \#TP/(\#TP + \#FN)$
False positive rate, *FPR*	$FPR = \#FP/(\#FP + \#TN)$
Positive predictive value, *PPV*	$PPV = \#TP/(\#TP + \#FP) = 1 - FDR$
False discovery rate, *FDR*	$FDR = \#FP/(\#TP + \#FP) = 1 - PPV$

Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

The R&D Process as a Series of Decision Steps

A. Simplified R&D process

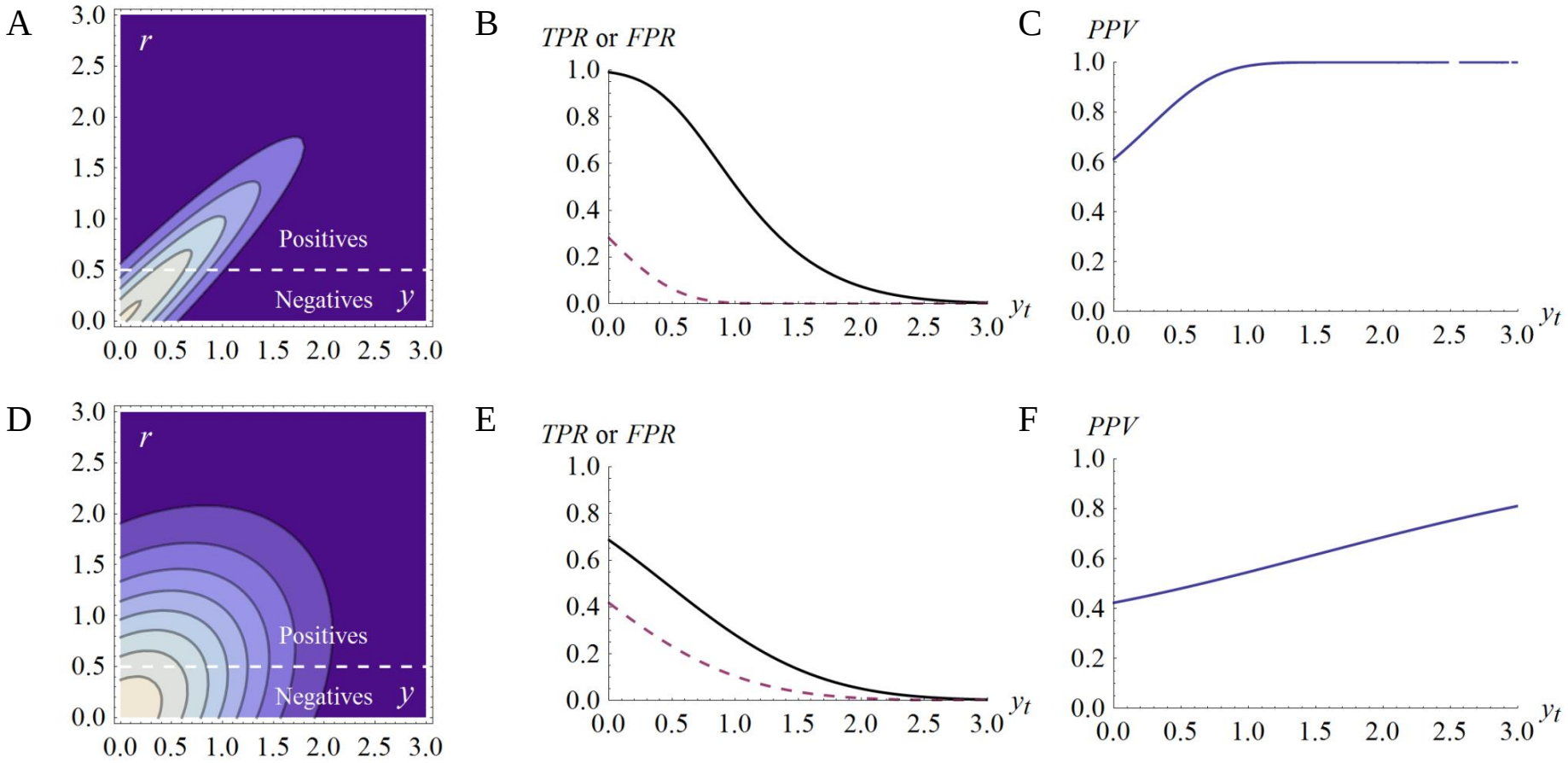


B. Correlation matrix for decision & reference variables

	X	Y	Z	R
X	1			
Y	$\rho_{X,Y}$	1		
Z	$\rho_{X,Z}$	$\rho_{Y,Z}$	1	
R	$\rho_{X,R}$	$\rho_{Y,R}$	$\rho_{Z,R}$	1

Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

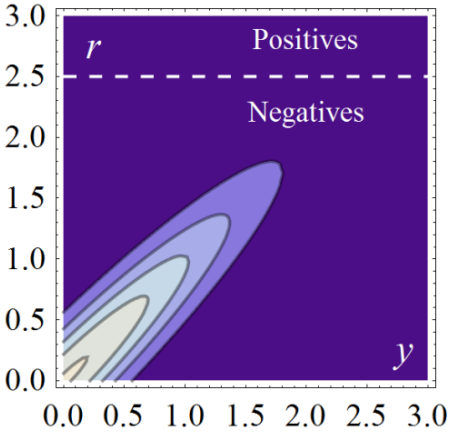
R&D Performance Varies with More vs Less Valid Models…



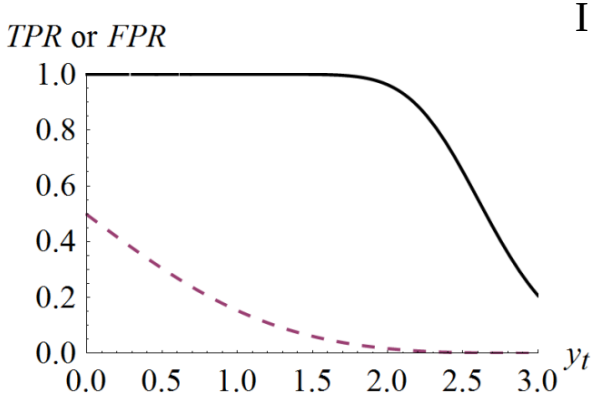
Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Less Valid Models do *Very* Badly When Positives are Rare

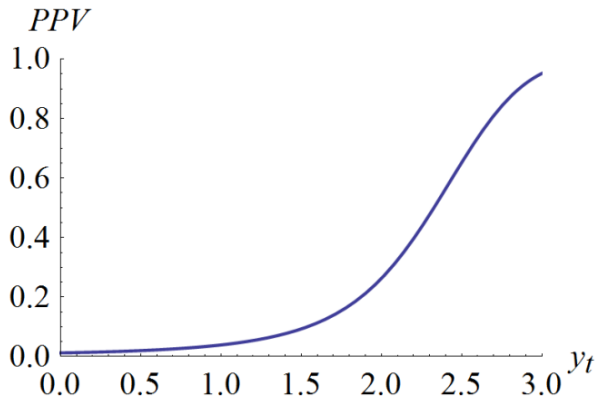
G



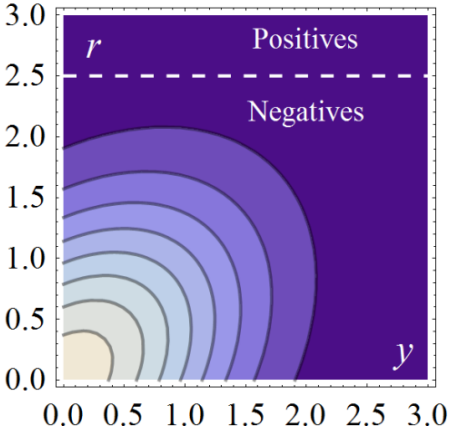
H



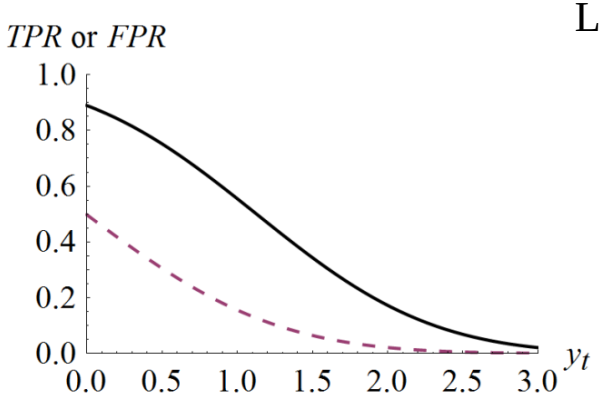
I



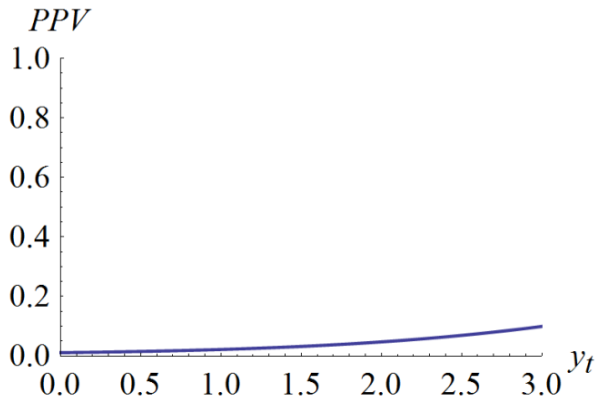
J



K



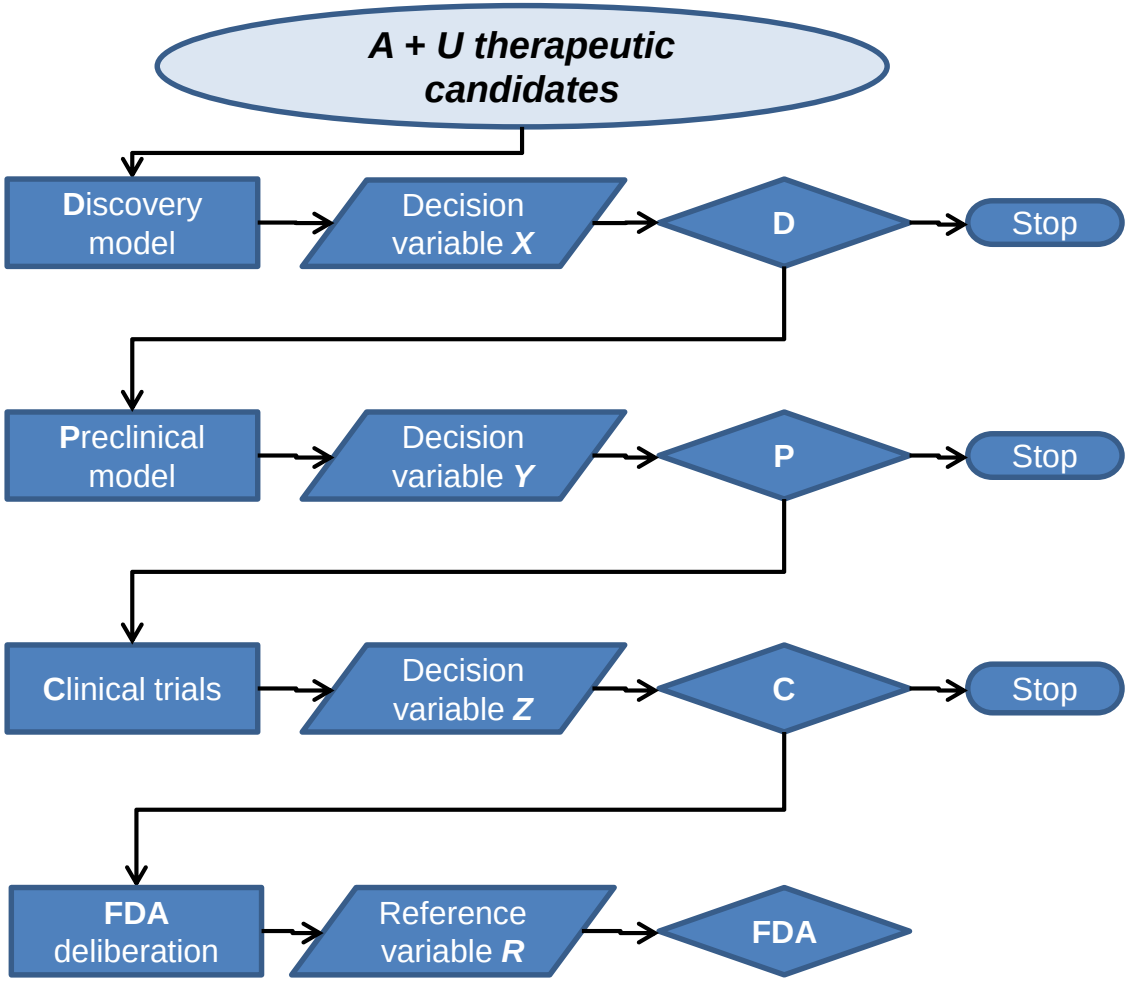
L



Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Modelling R&D as a Formal Search Process

A. Simplified R&D process



B. Correlation matrix for decision & reference variables

	X	Y	Z	R
X	1			
Y	$\rho_{X,Y}$	1		
Z	$\rho_{X,Z}$	$\rho_{Y,Z}$	1	
R	$\rho_{X,R}$	$\rho_{Y,R}$	$\rho_{Z,R}$	1

Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Compounding Effect of True and False Positive Rates

$$1)\ Q_{\text{start}} = \frac{A}{U}$$

$$2)\ Q_{\text{D}\rightarrow\text{P}} = \frac{A}{U} \times \frac{TPR_{\text{D}\rightarrow\text{FDA}}}{FPR_{\text{D}\rightarrow\text{FDA}}}$$

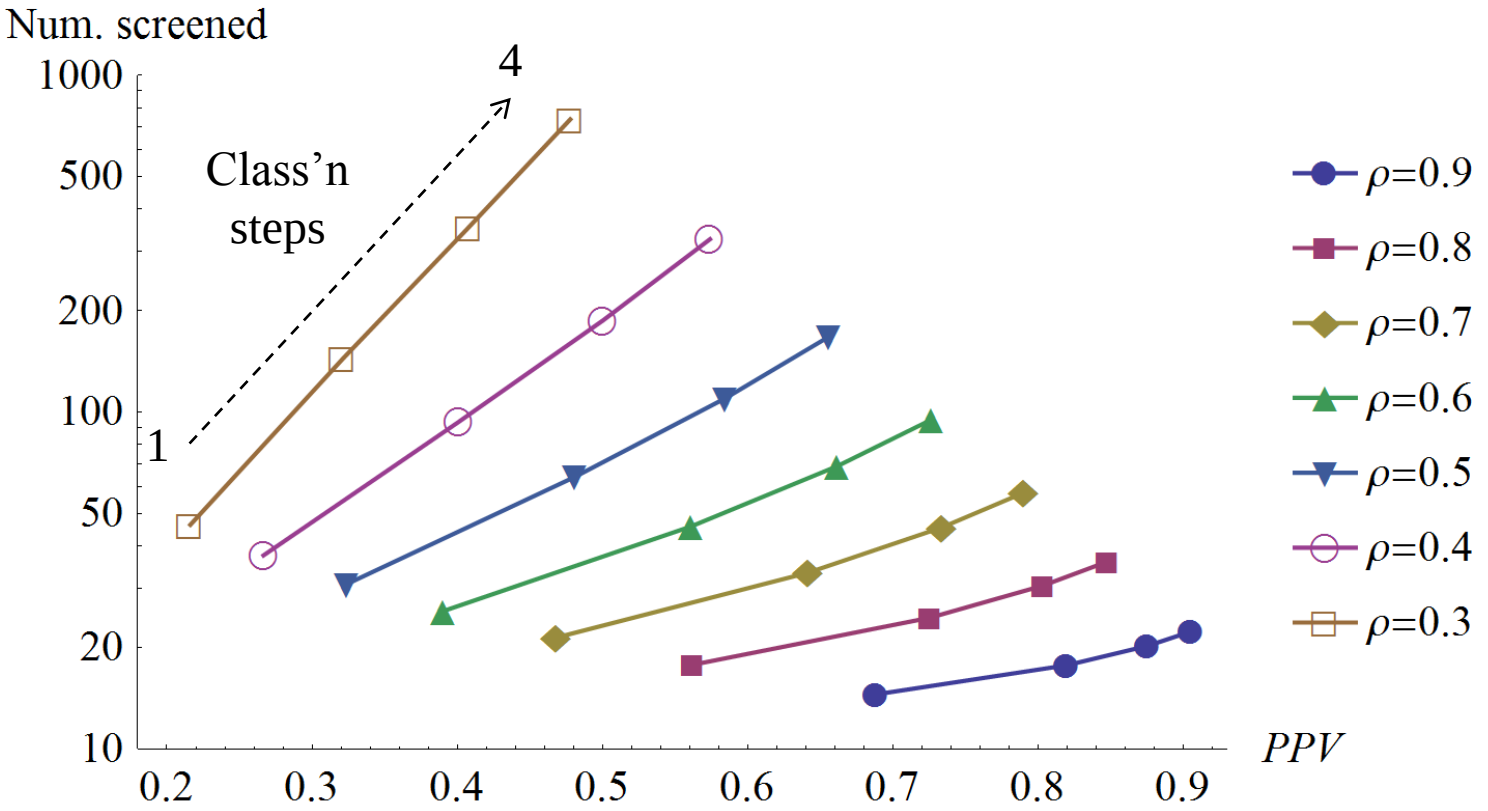
$$3)\ Q_{\text{P}\rightarrow\text{C}} = \frac{A}{U} \times \frac{TPR_{\text{D}\rightarrow\text{FDA}}}{FPR_{\text{D}\rightarrow\text{FDA}}} \times \frac{TPR_{\text{P}\rightarrow\text{FDA}}}{FPR_{\text{P}\rightarrow\text{FDA}}}$$

$$4)\ Q_{\text{C}\rightarrow\text{FDA}} = \frac{A}{U} \times \frac{TPR_{\text{D}\rightarrow\text{FDA}}}{FPR_{\text{D}\rightarrow\text{FDA}}} \times \frac{TPR_{\text{P}\rightarrow\text{FDA}}}{FPR_{\text{P}\rightarrow\text{FDA}}} \times \frac{TPR_{\text{C}\rightarrow\text{FDA}}}{FPR_{\text{C}\rightarrow\text{FDA}}}$$

Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Multiple Decision Steps: Quality Still Beats Quantity

PPV vs. candidates screened per *TP*

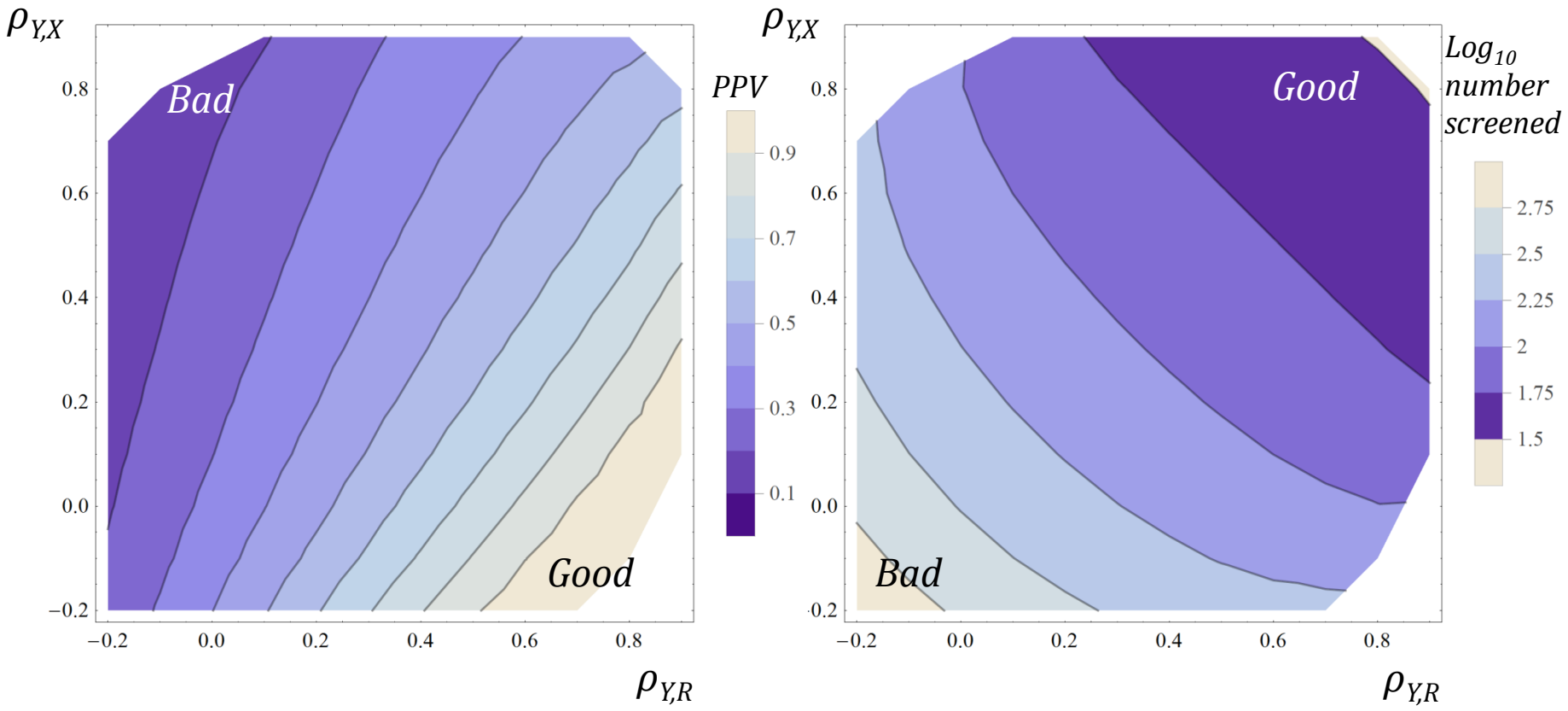


Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Multiple Decision Steps: Correlated with Other Models Does Not Mean "Validated"

A. *PPV*

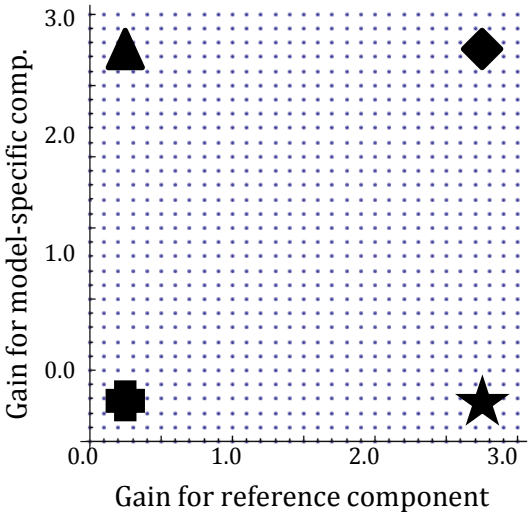
B. Candidates screened per *TP*



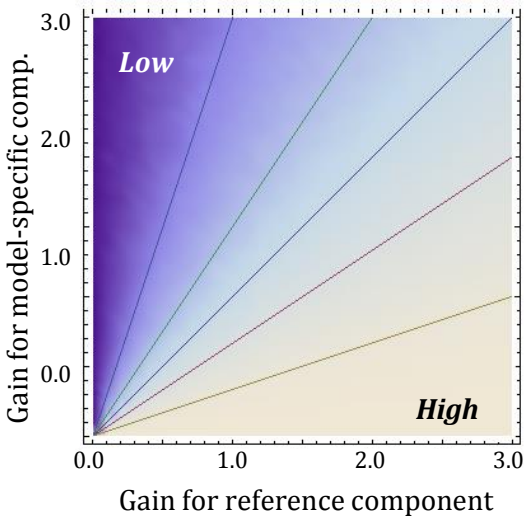
Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Reproducibility Declines Too if *Obviously* Valid Models are Retired

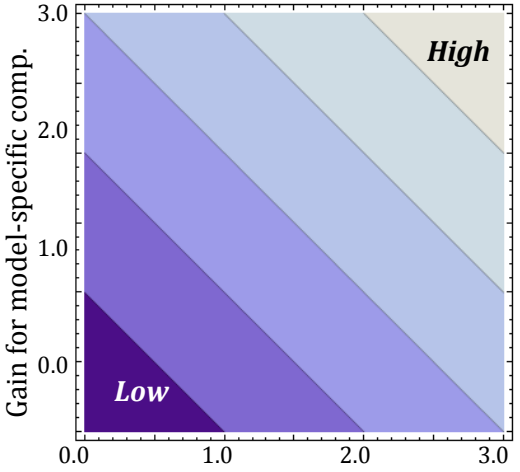
A. Signal has 2 components



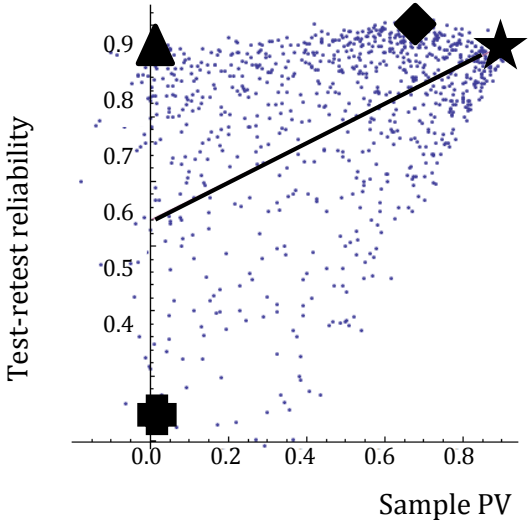
B. Predictive validity



C. Signal to noise ratio



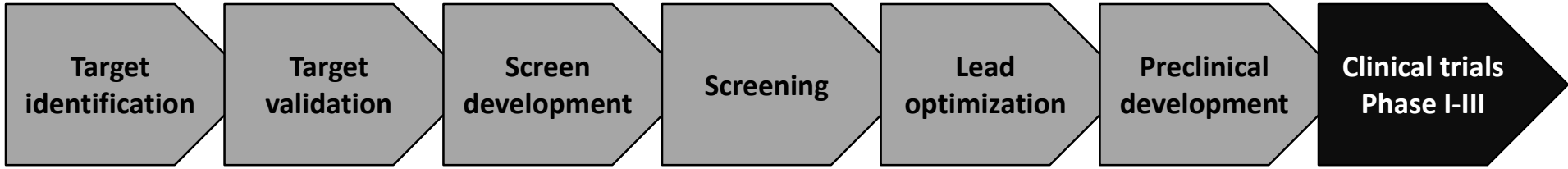
D. Model performance



Scannell & Bosley (2016) *When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis.* PLoS One. http://dx.doi.org/10.1371/journal.pone.0147215

Failure of Genomics and HTS to Deliver New Antibiotics

GSK Experience from 1995 to 2002



First bacterial genome sequenced in '95

Several major pathogens sequenced over next few years

> 350 candidate target genes identified from comparative sequence analysis

127 "essential" genes identified

Prioritised those with close homologues across variety of bacteria but without close homologues in man

67 High throughput screening campaigns conducted

Each involved testing gene products against up to ~500,000 compounds from the GSK compound libraries

Only 16 "hits" across all targets

16 "hits" led to only 5 "leads" with potentially attractive drug-like properties

However, leads had only narrow spectrum activity

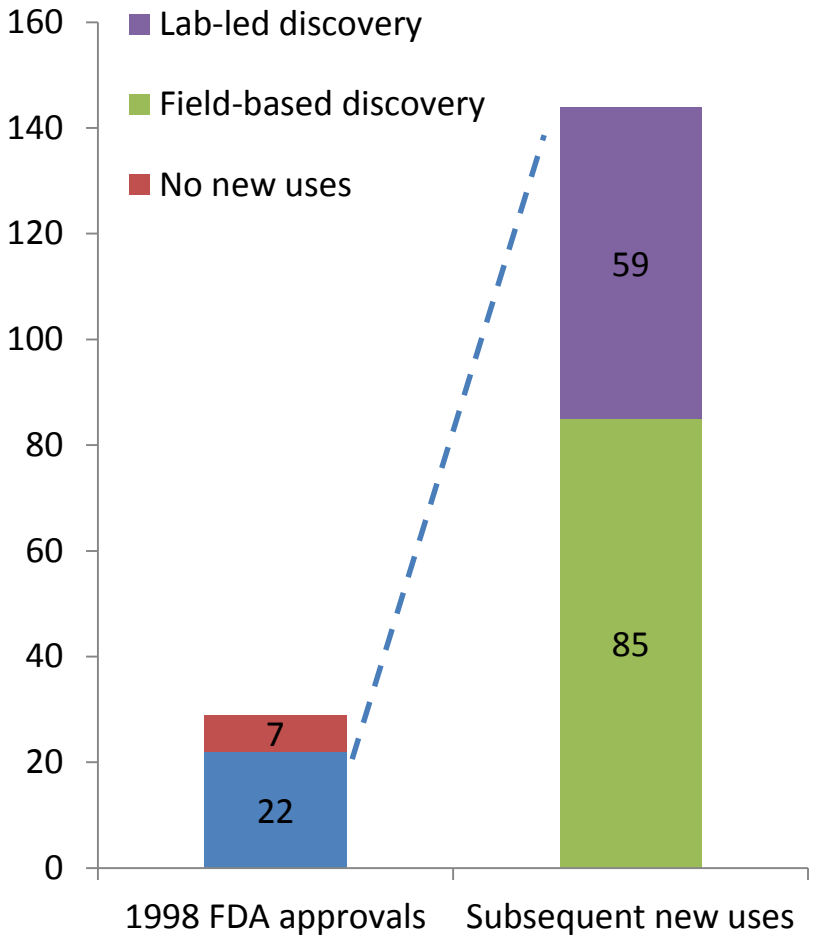
No clinical trials initiated

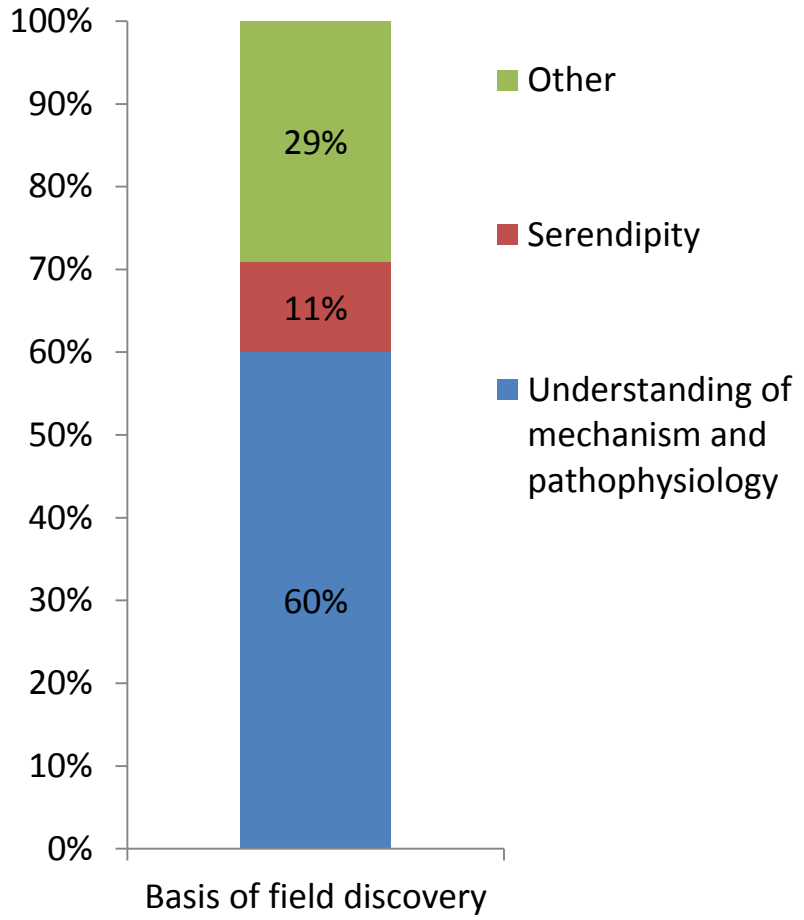
Source: Payne, D. J., Gwynn, M. N., Holmes, D. J., & Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nature Reviews. Drug Discovery, 6, 29–40. doi:10.1038/nrd2201.

See also: Silver, L. L. (2011). Challenges of antibacterial discovery. Clinical Microbiology Reviews, 24, 71–109. doi:10.1111/j.1749-6632.2010.05828.x

Note, this was not just a GSK problem. GSK has simply published its experience. According to Payne et al. (2007): *"GSK was not the only company that had difficulty finding antibacterial leads from HTS. A review of the literature between 1996 and 2004 shows that >125 antibacterial screens on 60 different antibacterial targets were run by 34 different companies. That none of these screens resulted in credible development candidates is clear from the lack of novel mechanism molecules in the industrial antibacterial pipeline."*

Field-based Discovery is Important: Low Throughput but High Predictive Validity





Source: DeMonaco HJ, Ayfer A, Von Hippel, E. (2006). The Major Role of Clinicians in the Discovery of Off-Label Drug Therapies. Pharmacotherapy, 26(3), 323–332.