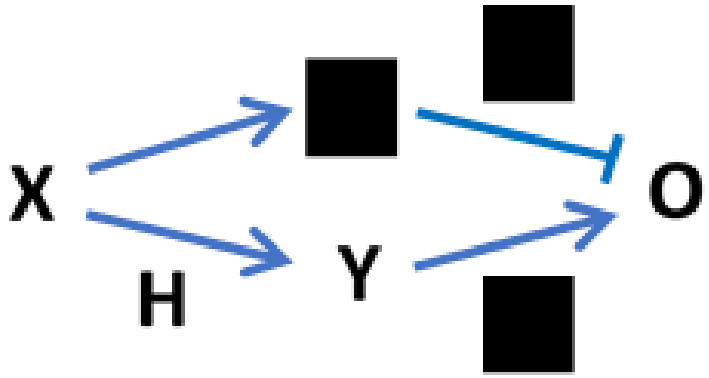
Predicting the Results of Medical Interventions:

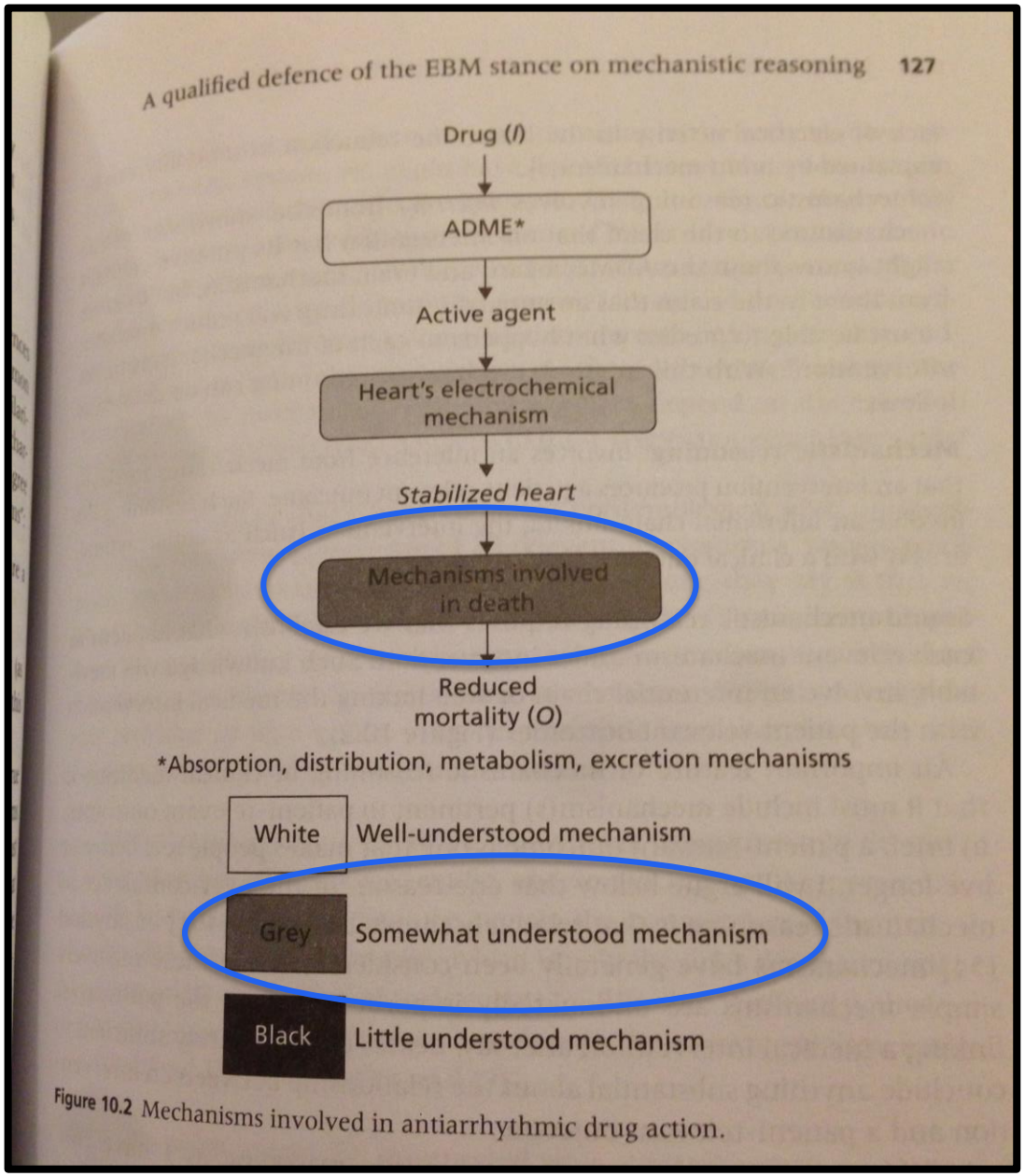


When Mechanistic Models Misfire

Jonathan Fuller
University of Toronto

Please do not cite or distribute without the author's permission:

Jonathan Fuller
jonathan.fuller@mail.utoronto.ca

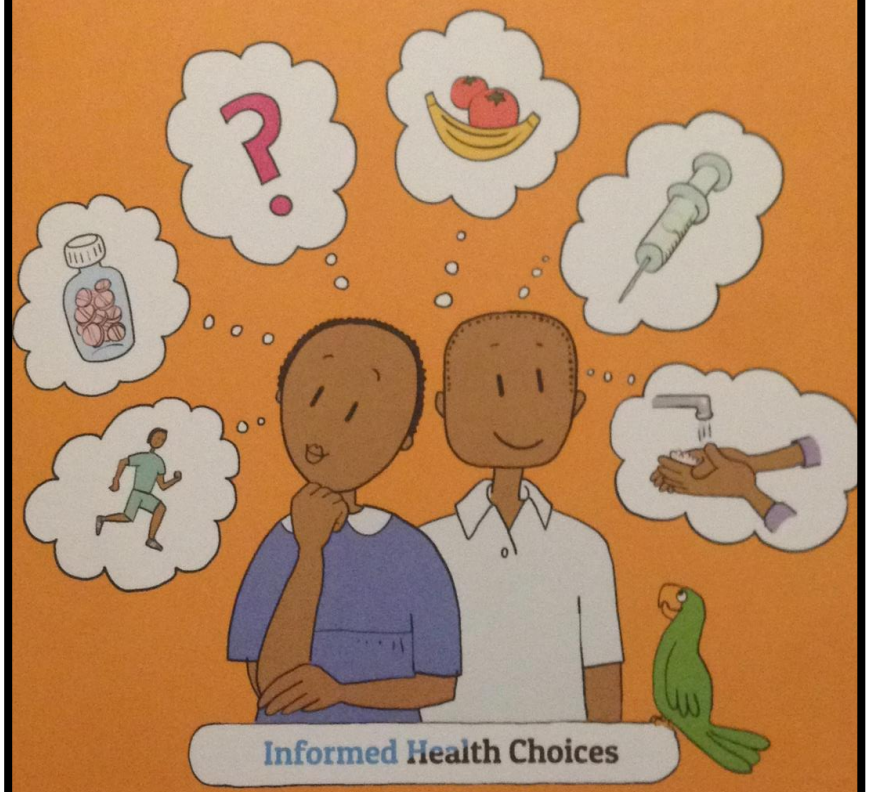


EBM

Howick (2011a)

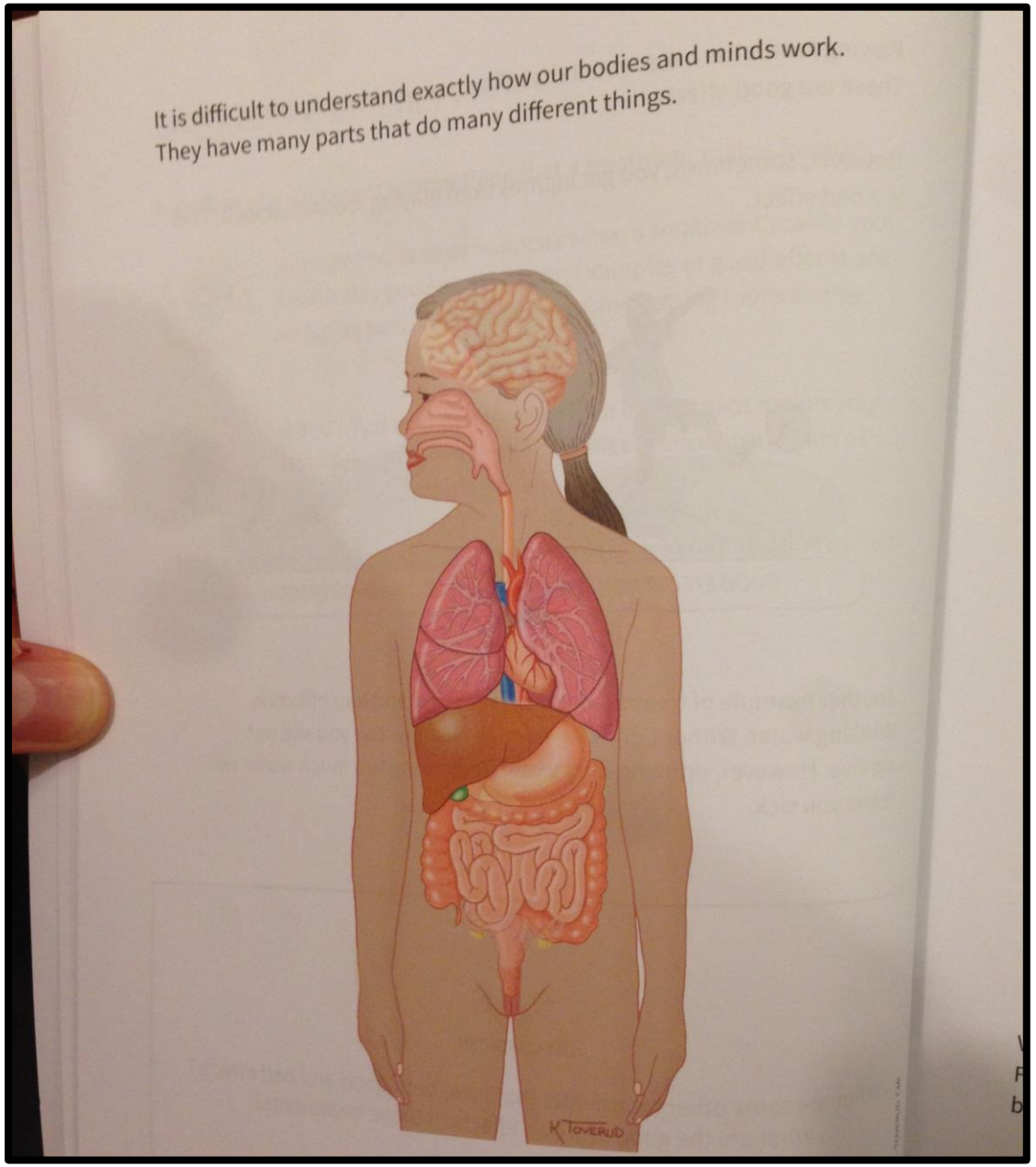
A health science book for primary school children

The Health Choices Book:
Learning to think carefully about treatments



Informed Health Choices

"It is difficult to understand exactly how our bodies and minds work. They have many parts that do many different things."



EBM+

Table 3 The GRADE system advocated by NICE (2009)

Criteria for assigning grade of evidence	
Type of evidence	Randomized trial = high
	Observational study = low
	Any other evidence = very low
Decrease grade if	Serious or very serious limitation to study quality
	Important inconsistency
	Some or major uncertainty about directness
	Imprecise or sparse data
	High probability of reporting bias
Increase grade if	Strong evidence of association—significant relative risk of >2 (<0.5) based on consistent evidence from two or more observational studies, with no plausible confounders (+1)
	Very strong evidence of association—significant relative risk of >5 (<0.2) based on direct evidence with no major threats to validity (+2)
	Evidence of a dose response gradient (+1)
	All plausible confounders would have reduced the effect (+1)
Range	High quality evidence
	Moderate quality evidence
	Low quality evidence
	Very low quality evidence

Evidence of correlation

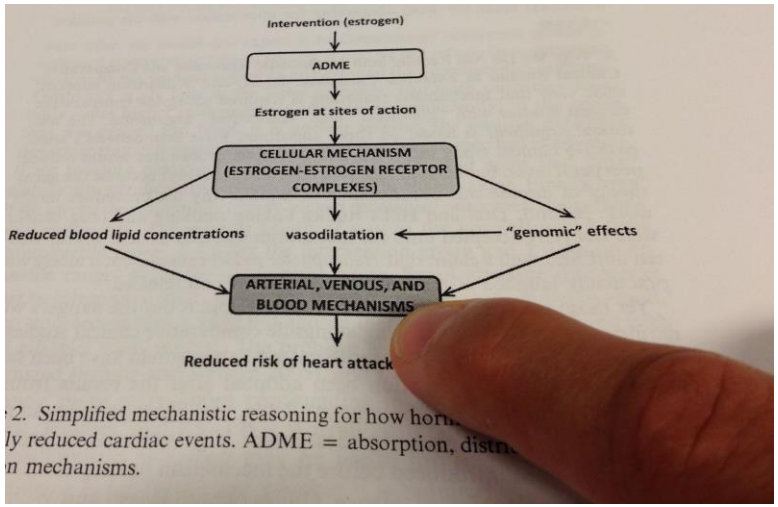
Table 6 Grading evidence of a mechanism

Pluses	Minuses
Each independent method that confirms a feature	Each independent method that fails to confirm—or, worse, disconfirms—a feature
Each independent research group that confirms a feature	Each independent research group that fails to confirm—or, worse, disconfirms—a feature
Larger proportion of features found	Smaller proportion of features found
Analogous mechanisms known	The analogy is a weak one, or, worse, analogous situations exhibit no such mechanism
Robust, reproducible across a wide range of conditions	Fragile, not reproducible in slightly varying conditions

Evidence of mechanisms

Clarke et al. (2014)

What could possibly go wrong?



2. Simplified mechanistic reasoning for how horm
ly reduced cardiac events. ADME = absorption, distri
n mechanisms.

Mechanistic reasoning?

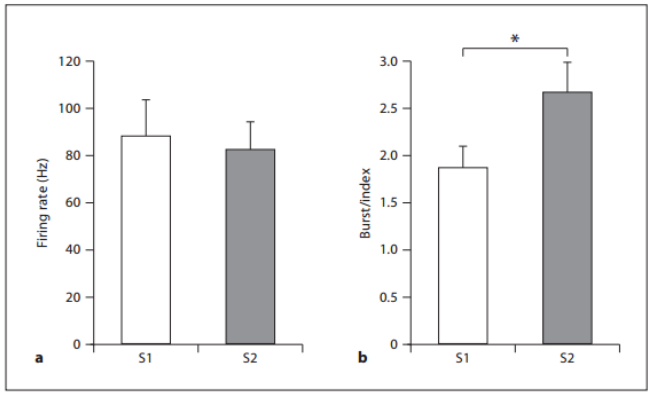


Fig. 2. Neuronal firing rates and bursting in the internal segment of the GPi. **a** Single-unit firing rates are averaged within S1 (intact putamen, n = 8) and S2 (degenerated putamen, n = 7). The mean firing rate did not differ significantly between sides. **b** Burst index values for single units are averaged within S1 (n = 8) and S2 (n = 7). The average burst index was significantly greater within S2 (* $p < 0.001$), suggesting a greater occurrence of bursting neurons.

Evidence of mechanisms?

Mechanisms misbehaving??

Main Objective:

Say some useful things (for the EBM+ project).

Main Contention:

To evaluate the 'quality' of mechanistic predictions (and other mechanistic inferences), we must assess the evidence for our mechanistic model *and* our model's correctness, completeness and abstractness.

Guiding Questions:

What concepts do we need?

Why do mechanistic predictions misfire?

What should we assess in order to evaluate the quality of mechanistic predictions?

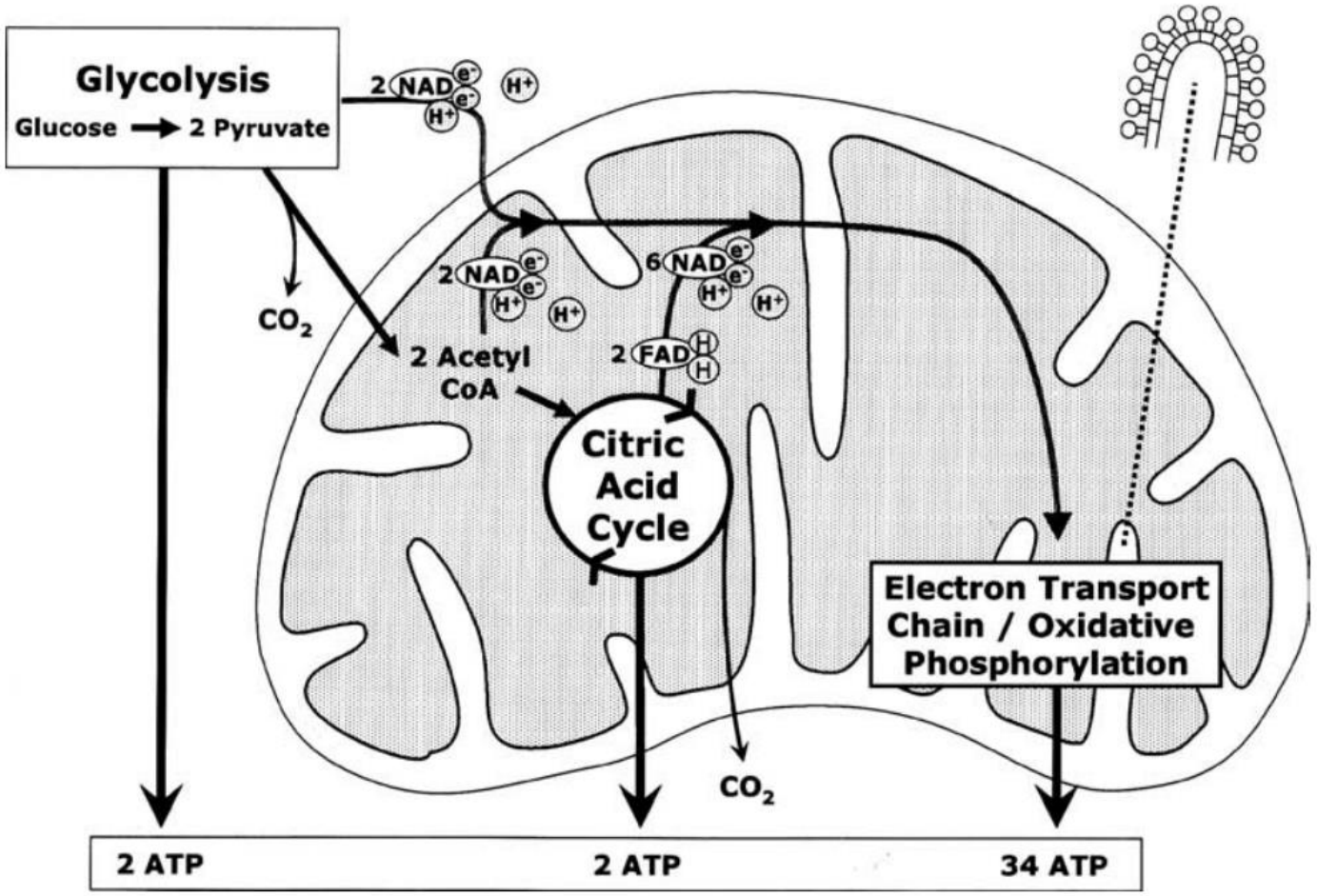
Guiding Questions:

What concepts do we need?

Why do mechanistic predictions misfire?

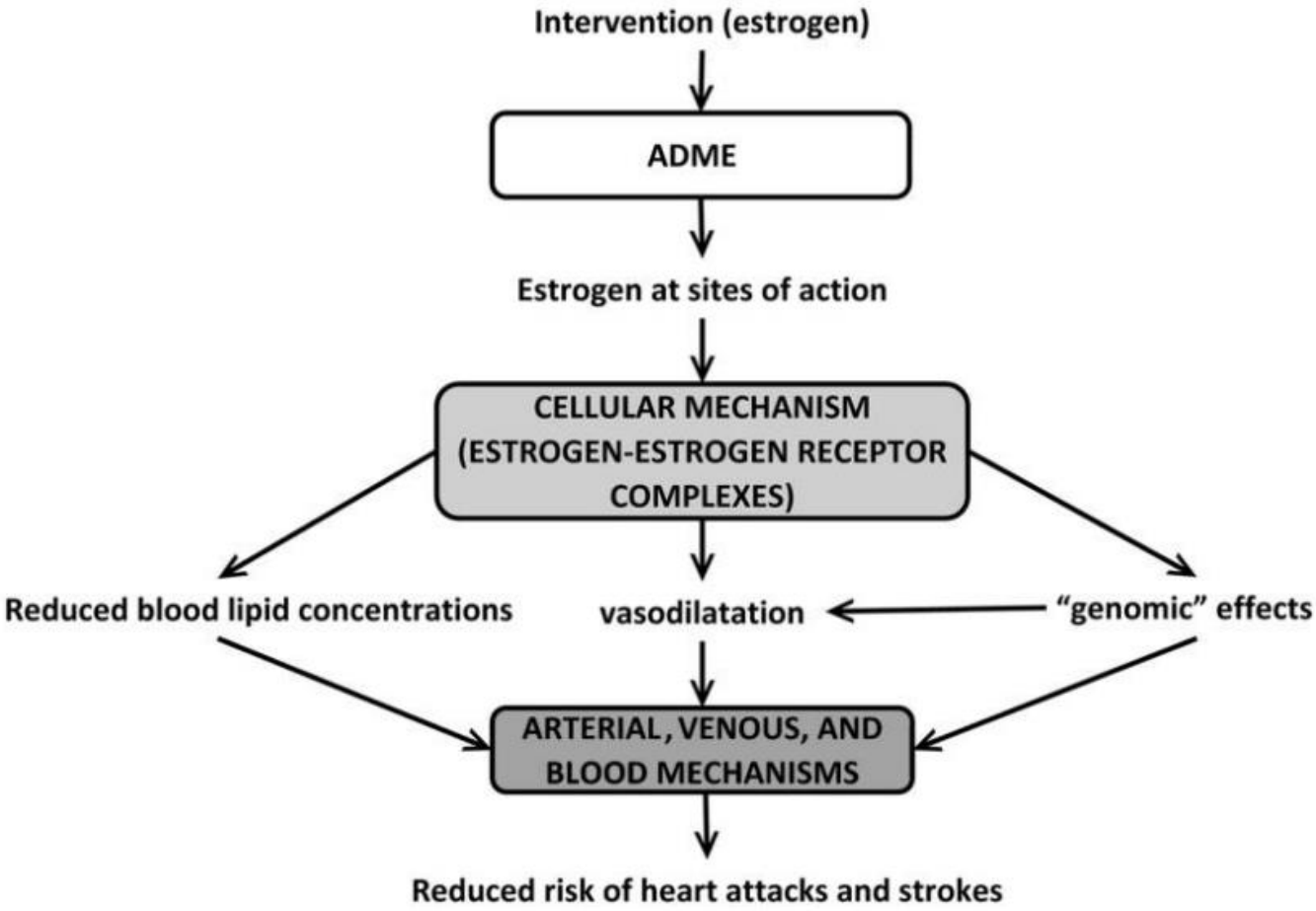
What should we assess in order to evaluate the quality of mechanistic predictions?

Mechanistic model



Bechtel & Abrahamson (2005)

Mechanistic model

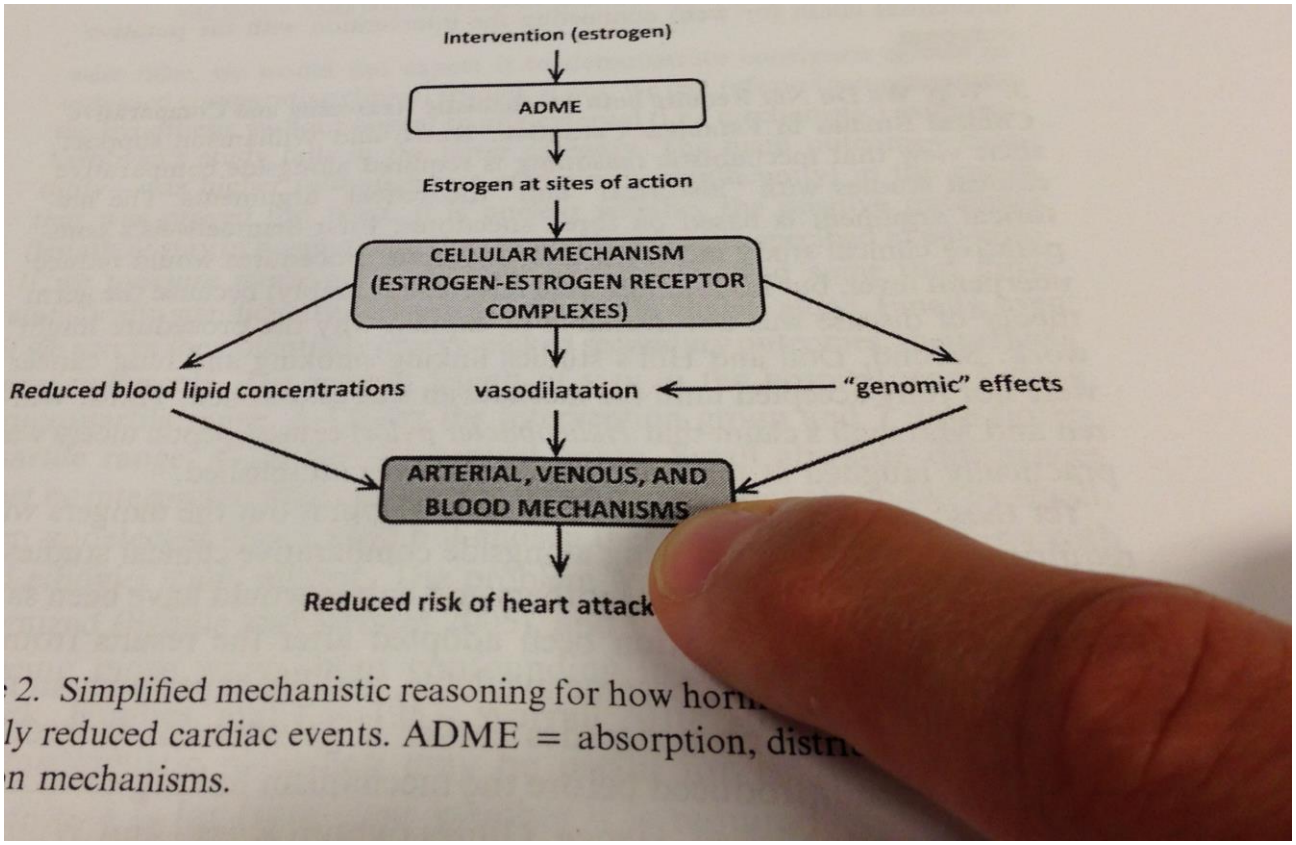


Howick (2011b)

Mechanistic model

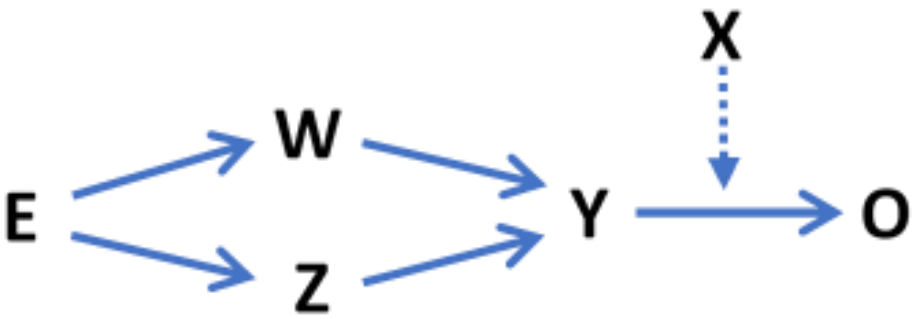
Reasoning through a mechanistic model

cf. ‘mechanistic reasoning’ (Howick 2011a,b)



Model of an Intervention Mechanism

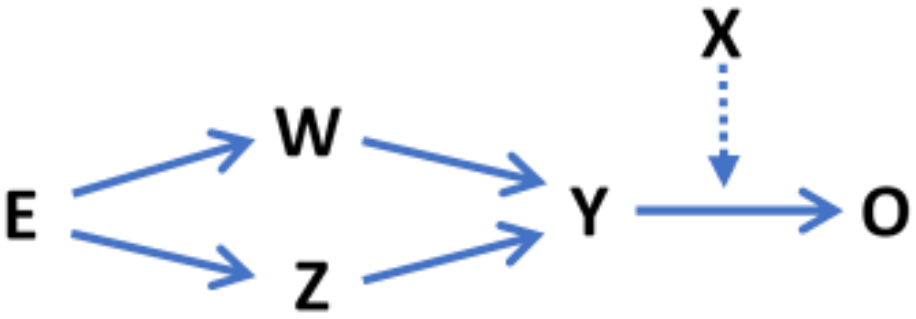
Model of an Intervention on a Mechanism



X: intervention

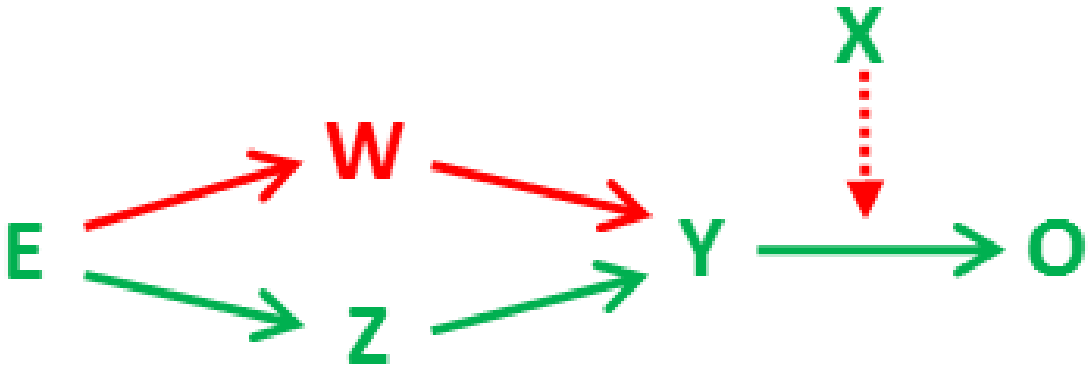
Mechanistic Prediction

Given X, predict that X will cause O.



Given E and X, predict that X will prevent O.

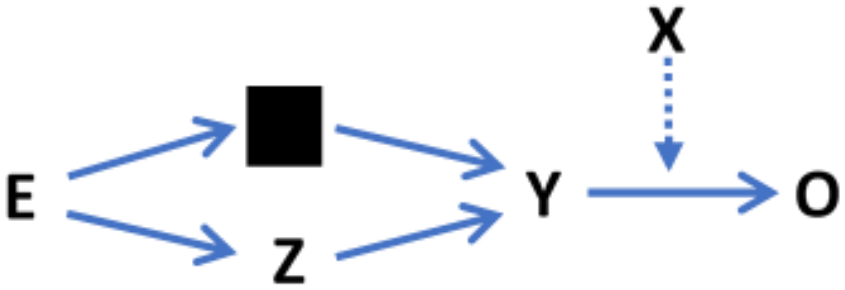
Properties of mechanistic models: correctness



Green: correct

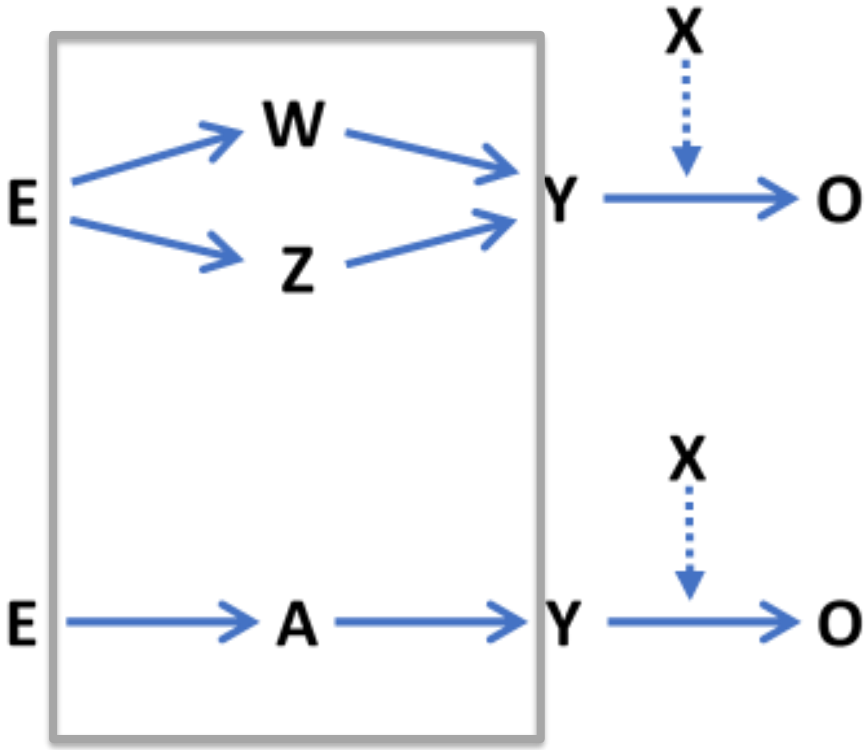
Red: incorrect

Properties of mechanistic models: completeness



Black box: incompleteness

Properties of mechanistic models: abstractness



(A represents W/Z)

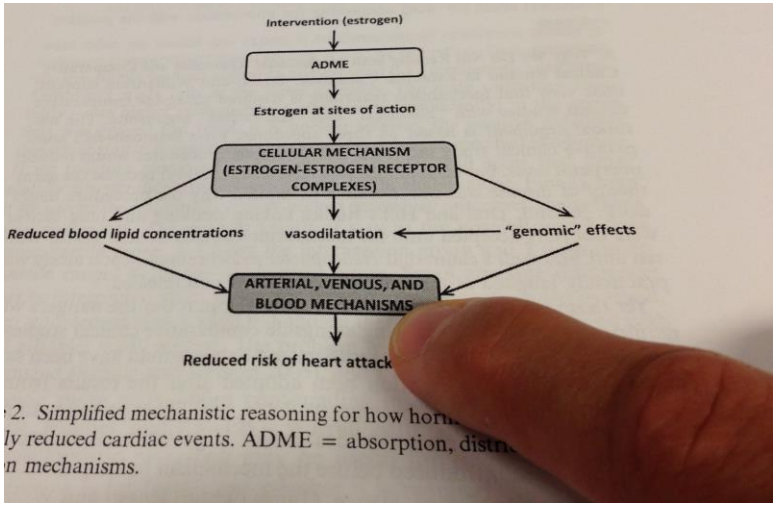
Guiding Questions:

What concepts do we need?

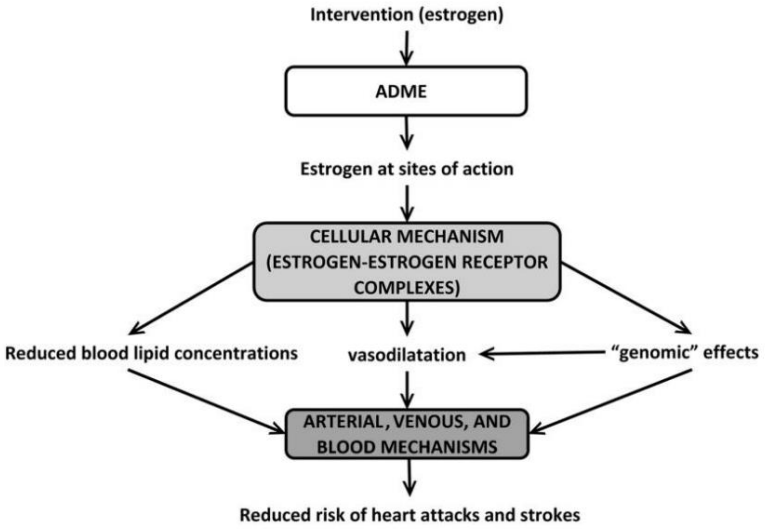
Why do mechanistic predictions misfire?

What should we assess in order to evaluate the quality of mechanistic predictions?

Whose to blame for a wrong prediction?



Our mechanistic reasoning?



Our mechanistic model?

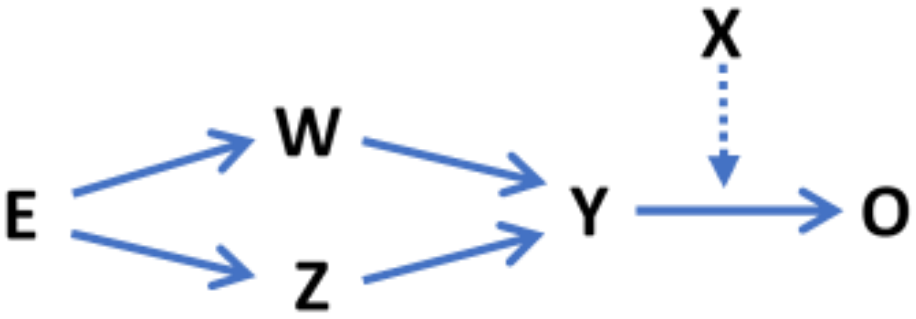
The mechanism?

Howick (2011a) – ‘problems with *mechanistic reasoning’*: unknown mechanisms, complex and stochastic mechanisms.

Clarke et al. (2014) – ‘limitations of *evidence of mechanisms*’: the Masking Problem, the Complexity Problem.

The reasoning is ‘valid’. But is it sound?

Given X, predict that X will cause O.



Given E and X, predict that X will prevent O.

Three ways that mechanistic models can mislead: incorrect, incomplete, too much/too little abstraction.

Incorrect?

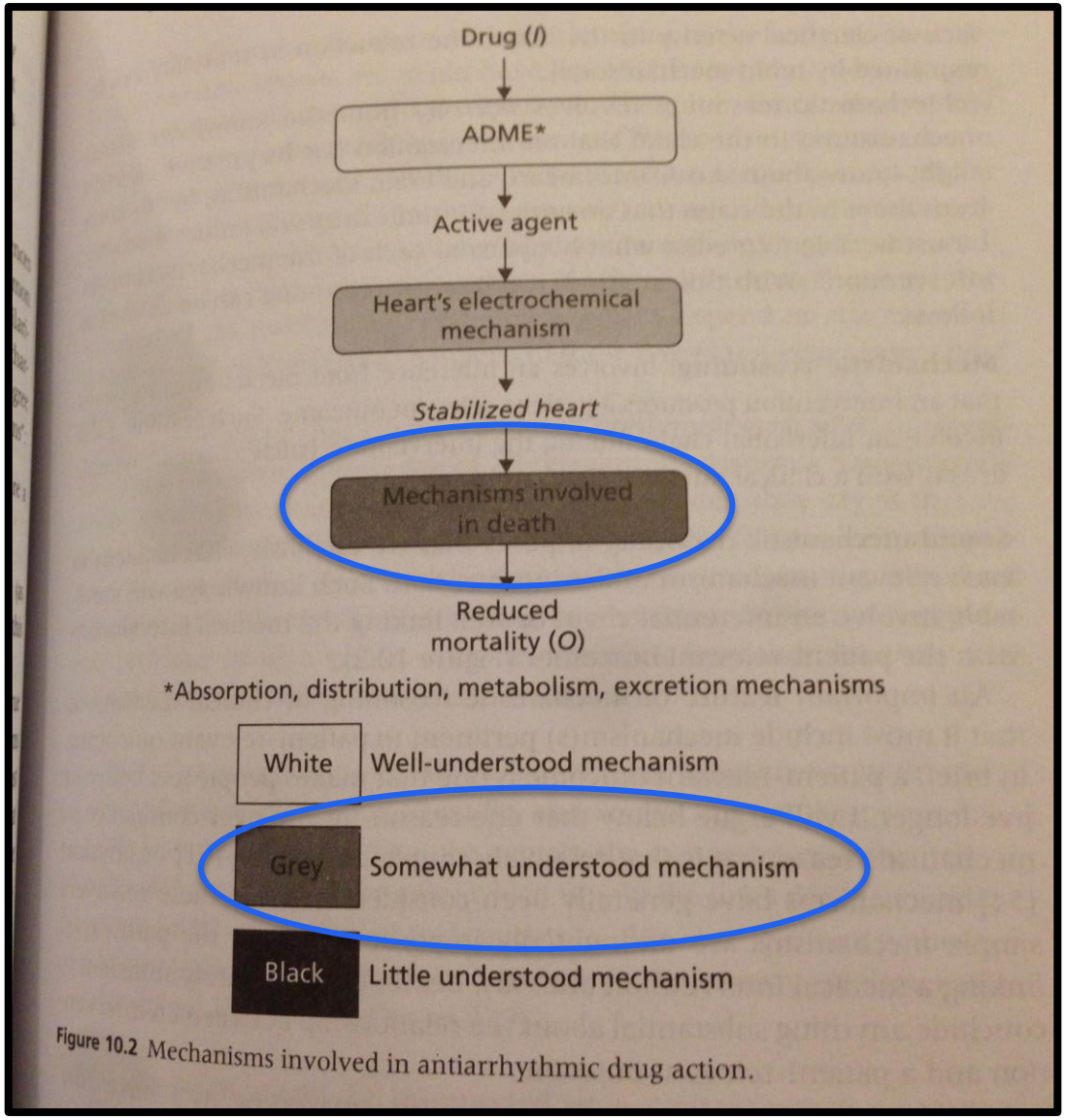


Figure 10.2 Mechanisms involved in antiarrhythmic drug action.

Howick (2011a)

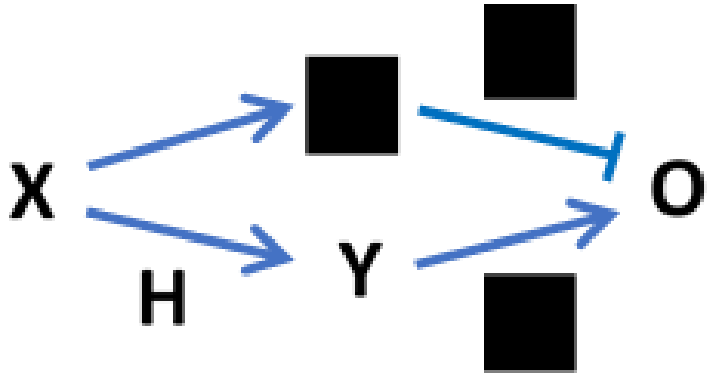
Incorrect?

X: antiarrhythmic drug
Y: stabilized heart rhythm
O: survival

Incomplete

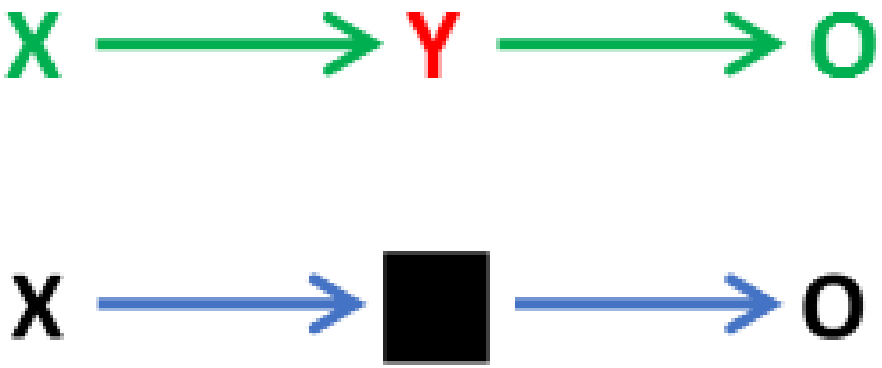
X: antiarrhythmic drug
Y: stabilized heart rhythm
O: survival

Incomplete



X: antiarrhythmic drug
Y: stabilized heart rhythm
O: survival
H: helping factors

Incorrect or incomplete but might still be good for predicting



Problems with incomplete mechanistic models?

Howick (2011a) – 'problems with *mechanistic reasoning'*: unknown mechanisms, complex and stochastic mechanisms.

Clarke et al. (2014) – 'limitations of *evidence of mechanisms*': the Masking Problem, the Complexity Problem.

Probabilistic and Quantitative Mechanistic Modeling

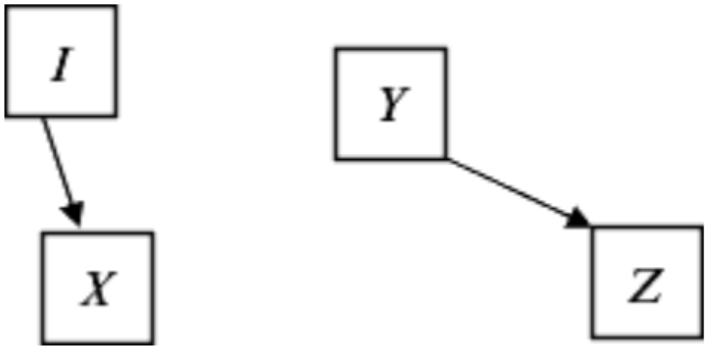


Figure 3.2 An ideal intervention

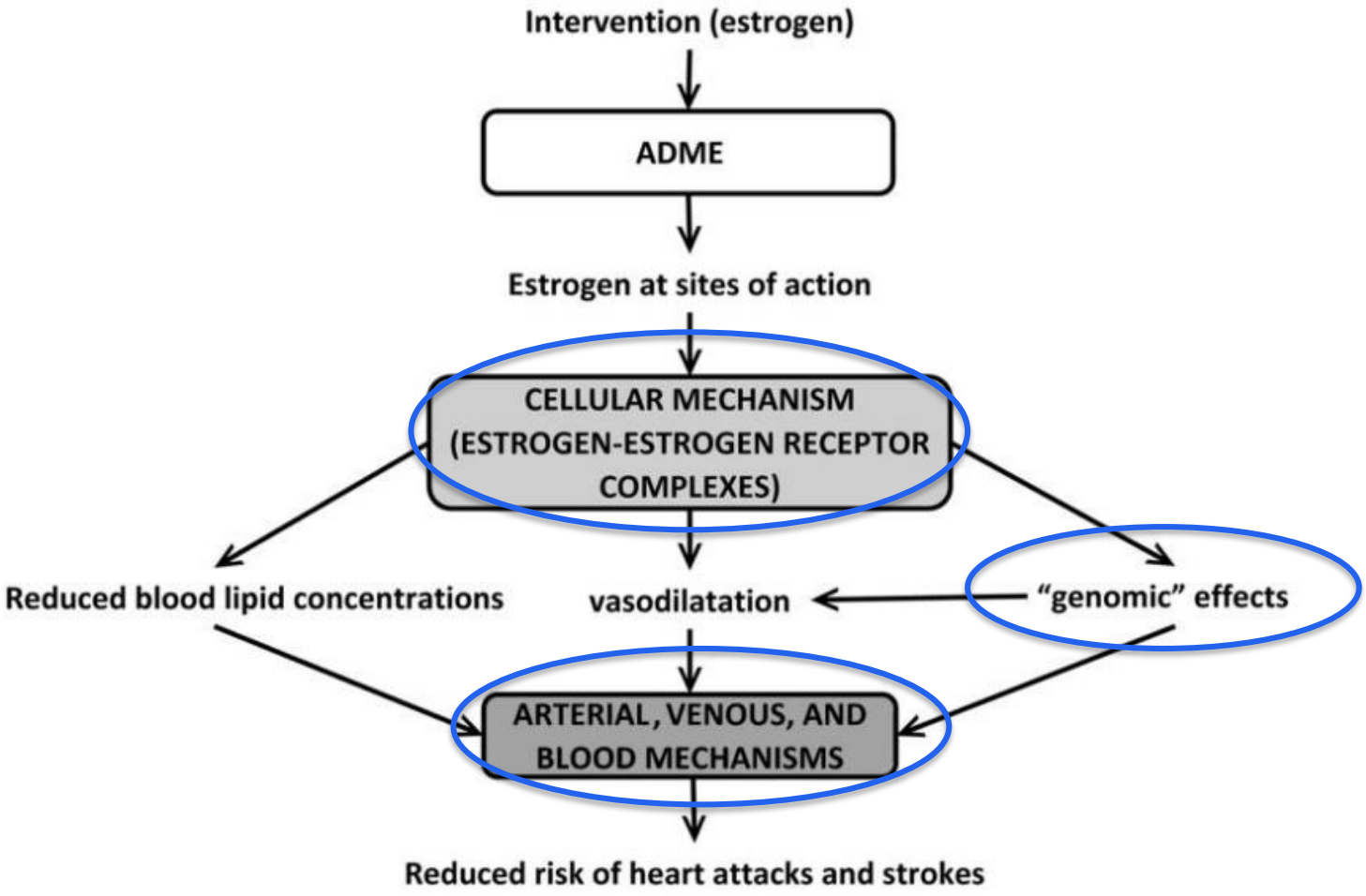
Causal Bayes nets interventions (Steel 2008)

Glucose: $(dY/dt) = (G - F3(X, Y) - F4(X, Y))/V$

That is, the change in extracellular glucose concentration with respect to time equals the rate of glucose injection minus the rate of liver accumulation of glucose minus the rate of tissue utilization of glucose, all divided by the volume of extracellular fluid.

Bolie Model of Insulin-Glucose Regulation (Thompson 2011)

Too Much Abstraction?



Howick (2011b)

Not Enough Abstraction?

educating mothers produces childhood nutrition

vs.

educating food buyers and providers produces childhood nutrition

Cartwright (2012)

Guiding Questions:

What concepts do we need?

Why do mechanistic predictions misfire?

What should we assess in order to evaluate the quality of mechanistic predictions?

1. Our Mechanistic Model

Is the mechanistic model incorrect, incomplete, very abstract/not very abstract *in ways that matter for the mechanistic prediction?*

Will our mechanistic prediction misfire?

vs.

2. Our Evidence for the Mechanistic Model

Evidence tells us whether the model is incorrect, incomplete, or very abstract in the first place.

cf. “evidence of the details of a *specific* mechanism” (Clarke et al. 2014).

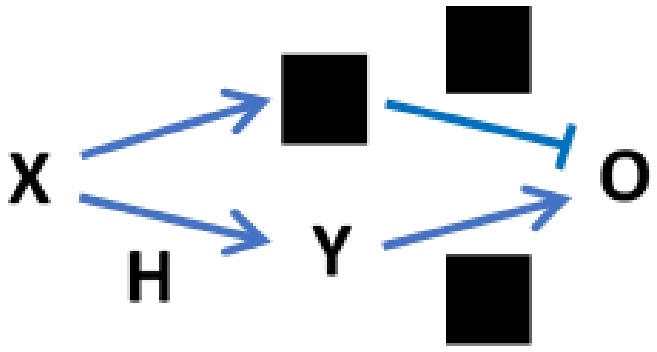
Main Contention:

To evaluate the 'quality' of mechanistic predictions (and other mechanistic inferences), we must assess the evidence for our mechanistic model *and* our model's correctness, completeness and abstractness.

Table 6 Grading evidence of a mechanism

Pluses	Minuses
Each independent method that confirms a feature	Each independent method that fails to confirm—or, worse, disconfirms—a feature
Each independent research group that confirms a feature	Each independent research group that fails to confirm—or, worse, disconfirms—a feature
Larger proportion of features found	Smaller proportion of features found
Analogous mechanisms known	The analogy is a weak one, or, worse, analogous situations exhibit no such mechanism
Robust, reproducible across a wide range of conditions	Fragile, not reproducible in slightly varying conditions

Evidence for mechanistic model



Mechanistic model

Acknowledgements

Support from: Canadian Institutes of Health Research (CIHR) and the McLaughlin Centre.