

Generating & Evaluating Evidence for Obesity Prevention

Alternatives and trade-offs

J. Michael Oakes, PhD

**McKnight Presidential Fellow
Associate Professor**

**Division of Epidemiology
University of Minnesota
*oakes007@umn.edu***

On being asked to talk on the principles of research, my first thought was to arise... and say, "Be careful", and to sit down.

J Cornfield 1959

Issues outlined in memo:

1. Do null results/studies mean no effect?
2. Success and failure in prevention trials?
3. Systems thinking?

1. Do null results/studies mean no effect?

Probably, but:

Effect heterogeneity?

Subjects actually treated/exposed?

Secular trends?

Measurement error?

Statistical power?

Of course, a small p-value doesn't mean Tx worked either!

Another thing...

Proper/useful inference from **observational designs** (ie, hypothetical interventions) is very difficult.

See epi HRT research

In terms of prevention, you should discount all claims about effects discovered in observational studies. Consider them hypotheses to be evaluated.

2. “Success” and “failure” in prevention trials?

Depends on goal(s) of study:

Tx development?

Recruitment?

Retention?

Implementation/Fidelity?

Null hypotheses ($d=0$)?

Estimators: ACE, ATT, ITT, LATE?

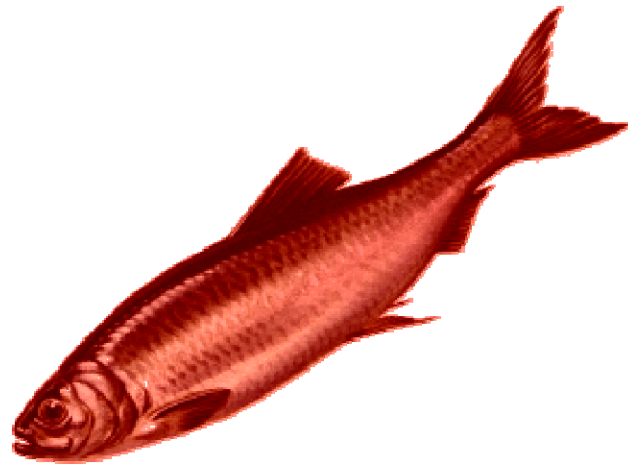
3. Systems thinking?

Probably a good thing...

One thing leads to another...

Feedback loops, cultural trends, etc

But



Consider:

Cohen, J. 1994. "The earth is round ($p < 0.05$)" *Amer Psychologist*. 49:997-1003

Poole, C. 2001. "Low P-values or narrow confidence intervals: which are more durable?" *Epidemiology* 12:291-4.

Smith, GC, and J. P. Pell. 2003. "Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized controlled trials." *BMJ* 327:1459-61.

Greenland, S. 1992. "Falsificationism and clinical trials." *Stat in Med* 11:1263-5.

Cochran, WG. 1965. "The Planning of Observational Studies of Human Populations." *J Royal Stat Society, Ser A* 128:243-265.

Hernan, MA, et al. 2008. "Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease." *Epidemiology* 19:766-79 (with discussion).

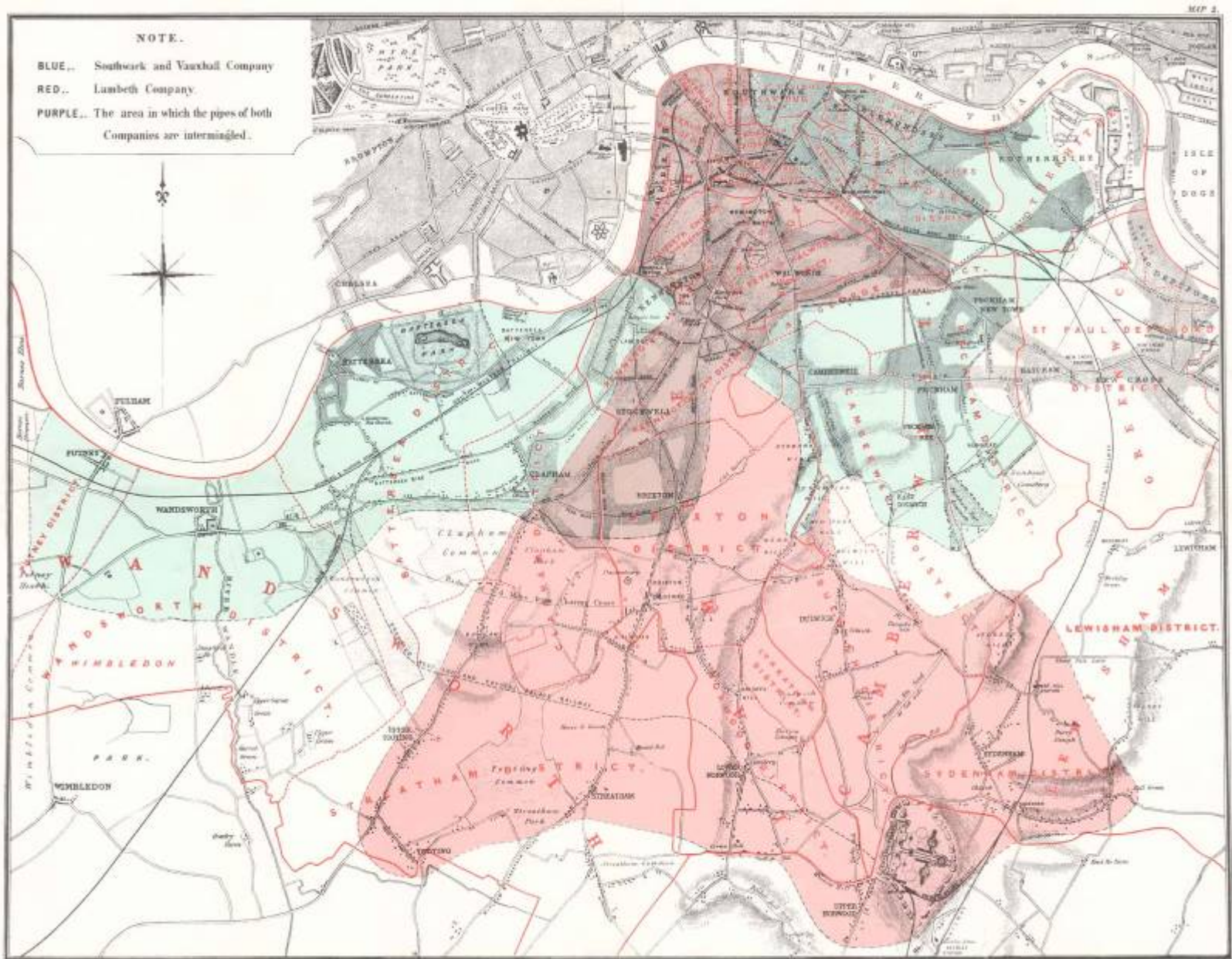
Consider these revolutionary studies too

No fancy statistical models here

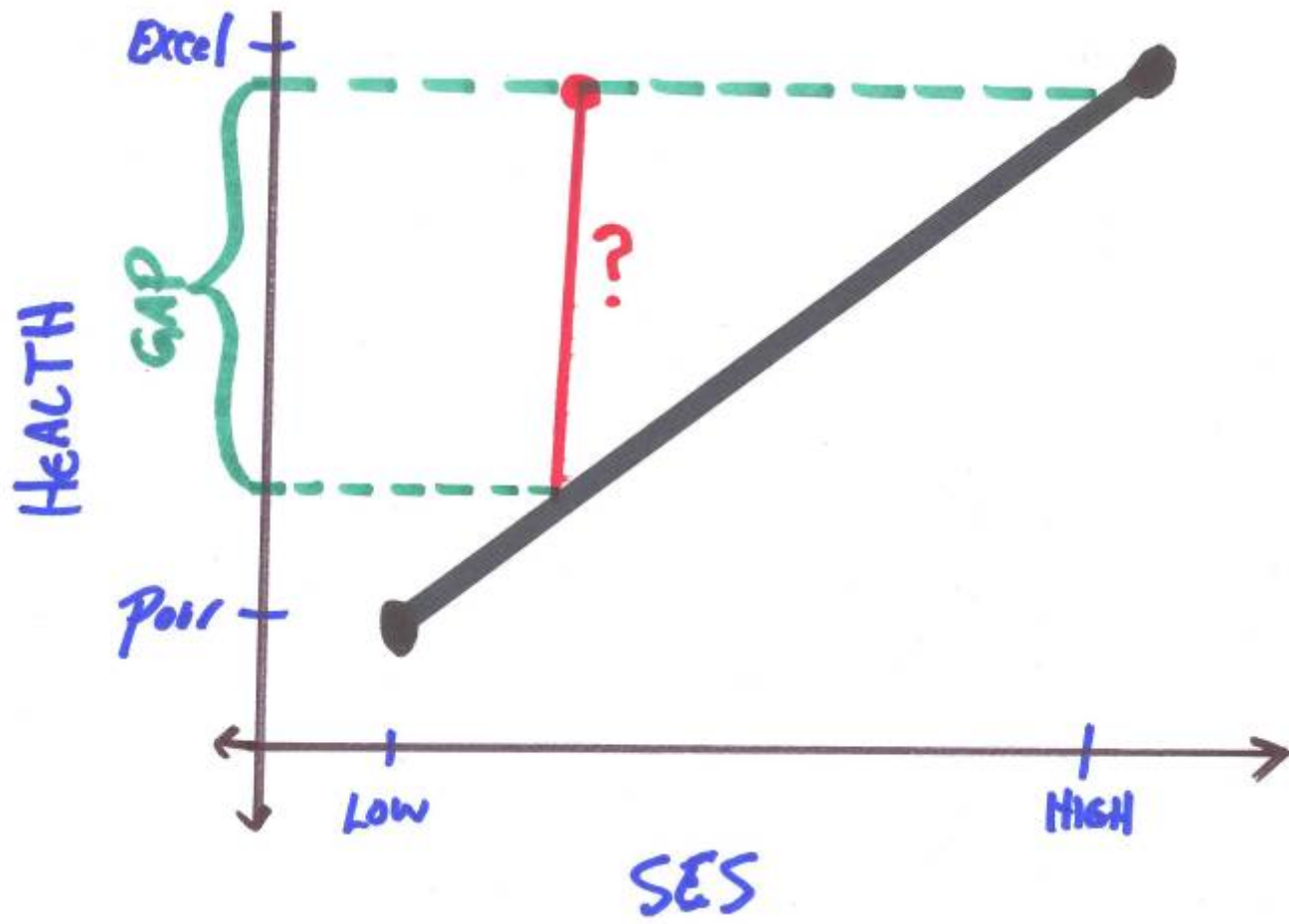
- Jenner (1796) + cow pox
- Semmelweis (1844) + puerperal fever*
- Snow (1854) + cholera*
- McKay (1901) + tooth decay*
- Goldberger (1914) + pellagra*

* Natural experiment

Other stuff...



Exchangeability in context effects



SES and health

Is exposure enough?

Consider Lubeck Germany 1929

- TB vaccination of 412 infants
- Oops, 251 with “live” TB bug
- 71 died, but 180 (71%) did not!

More food for thought...

1. Sample size
2. Response rates
3. Off-support inference
4. Inference and study design

1. The more subjects the better

**True in experiments.
But in observational studies:**

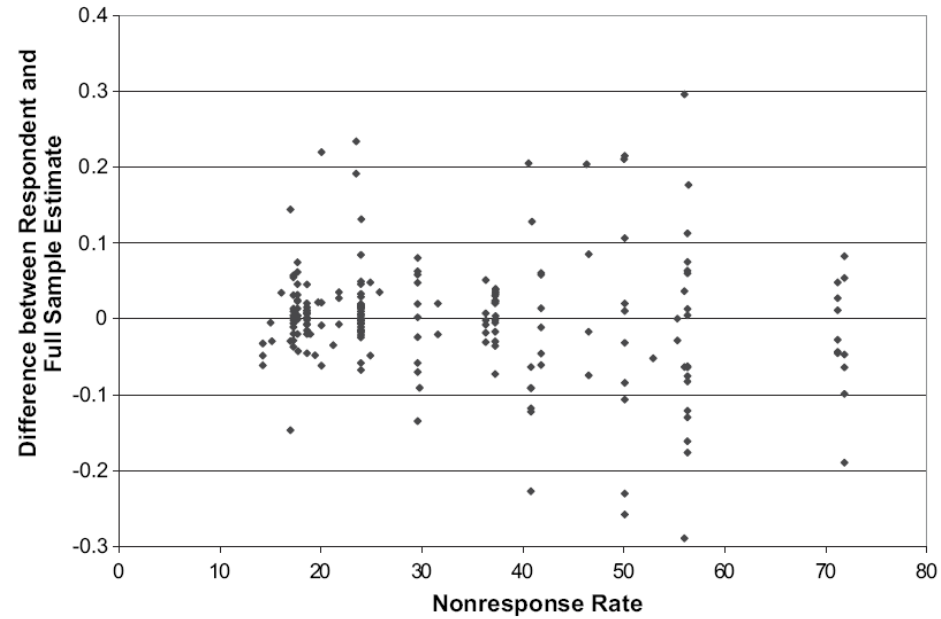
Reducing heterogeneity reduces *both*
sampling variability and sensitivity to hidden bias.

Increasing sample size increases precision but does little to reduce bias.

Even if smaller, homogeneous studies are better!

Or, beware of p-values from big fat national sample data.

2. Survey quality begins at 65%



Response bias and Nonresponse rates

within and between 23 studies (191 estimates)

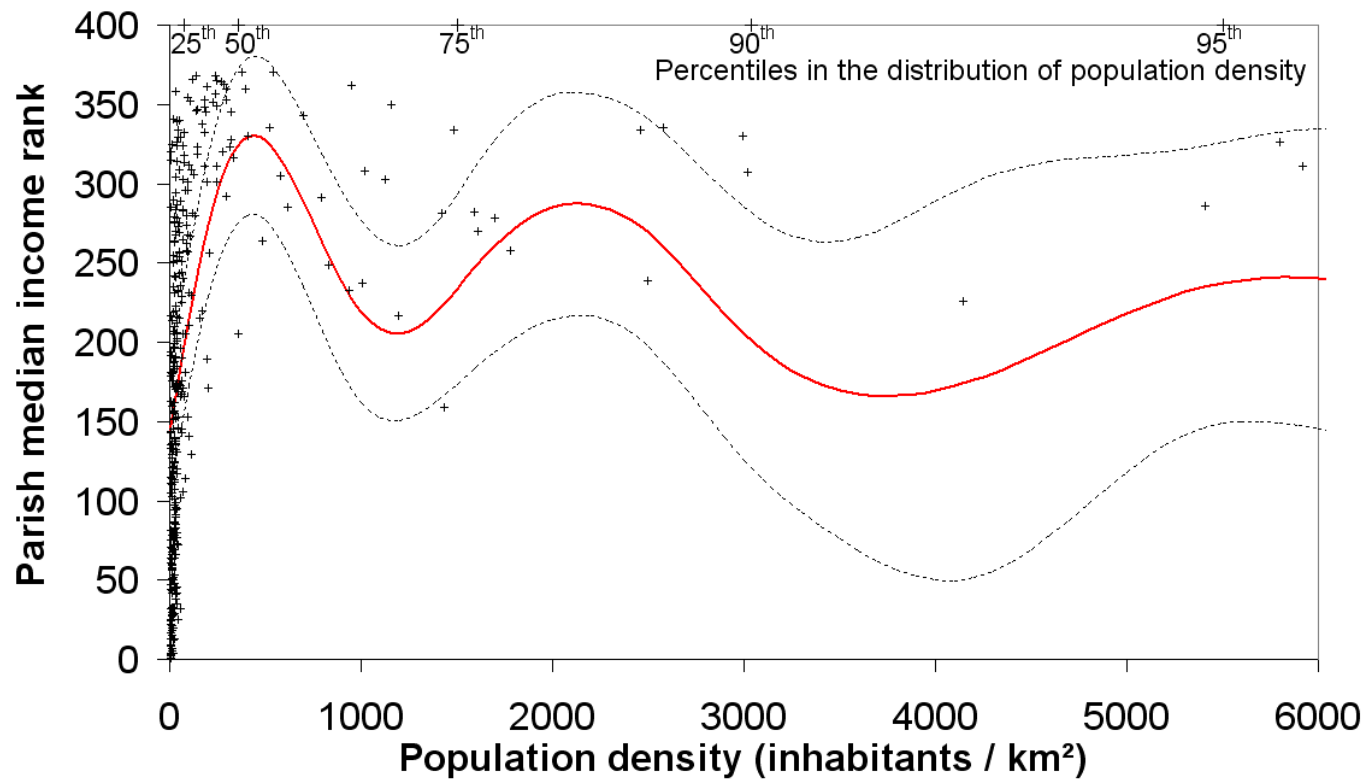
Survey response rates are not empirically useful
as they don't measure response bias,
which is central to sound inference.

Even if less, better quality data yield better results.

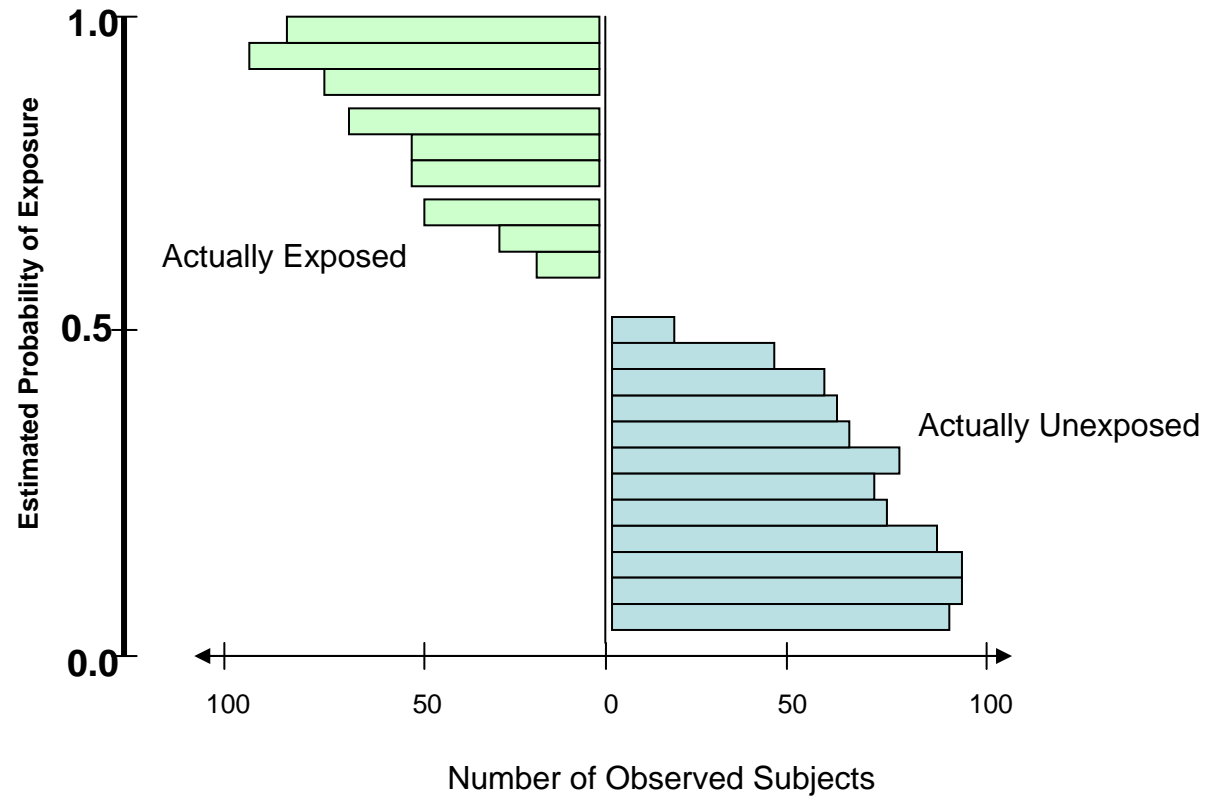
Or, junk in always yields junk out.

3. Structural confounding





		Density			
		1 st Quartile	2 nd Quartile	3 rd Quartile	4 th Quartile
SES	4 th Quartile	> 50	50	12	6
	3 rd Quartile	> 50	12	8	7
	2 nd Quartile	> 50	2	1	0
	1 st Quartile	> 50	0	0	0



Comparative inference is off-support of data and thus requires “heroic” modeling assumptions.

Source: Oakes, JM and PJ Johnson. 2006. "Propensity score matching methods for social epidemiology." Pp. 370-392 in *Methods in Social Epidemiology*, edited by Oakes and Kaufman. San Francisco: Jossey-Bass

See also – Johnson PJ. 2004. "The Effect of Neighborhood Poverty on American Indian Infant Death." PhD Dissertation, UMN
 Hearst MO. 2007. "The Effect of Racial Residential Segregation on Infant Death in the US." PhD Dissertation, UMN

'We can only evaluate sharply distinct treatments that could happen to anyone.'

Paul Rosenbaum (2002)

'If the differences between groups is large, the average value applied to each group with adjustment may represent "no man's land", a place where no actual observations exist. Given this scenario, the interpretation of the estimate becomes speculative rather than soundly based. Heroic modeling assumptions are required.'

William Cochran (1957)

4. Research designs

- Experiments (group or subject)
- Natural Experiments (group or subject)
- Quasi-Experiments (group or subject)
- Observational Studies
- Surveys

Experiments & Observational Studies

In experiments we theorize from **cause to effect**
(from known intervention to outcome)

If we intervene and randomize, we're in good shape.
But simply intervening (quasi-experiment) is also helpful.

Absent intervention,
we must theorize from **effect to cause**
which is not easy, especially in presence of confounding.
Results could be due to anything and what of unmeasured things?

Experimental Studies

If you analyze data from an experimental study you can stop at a “statistically significant” result.

But you must work to evaluate the actual fidelity of the study to the experimental ideal (eg, compliance)

90% of the work in an experimental study lies with preparation and design. This is the key.

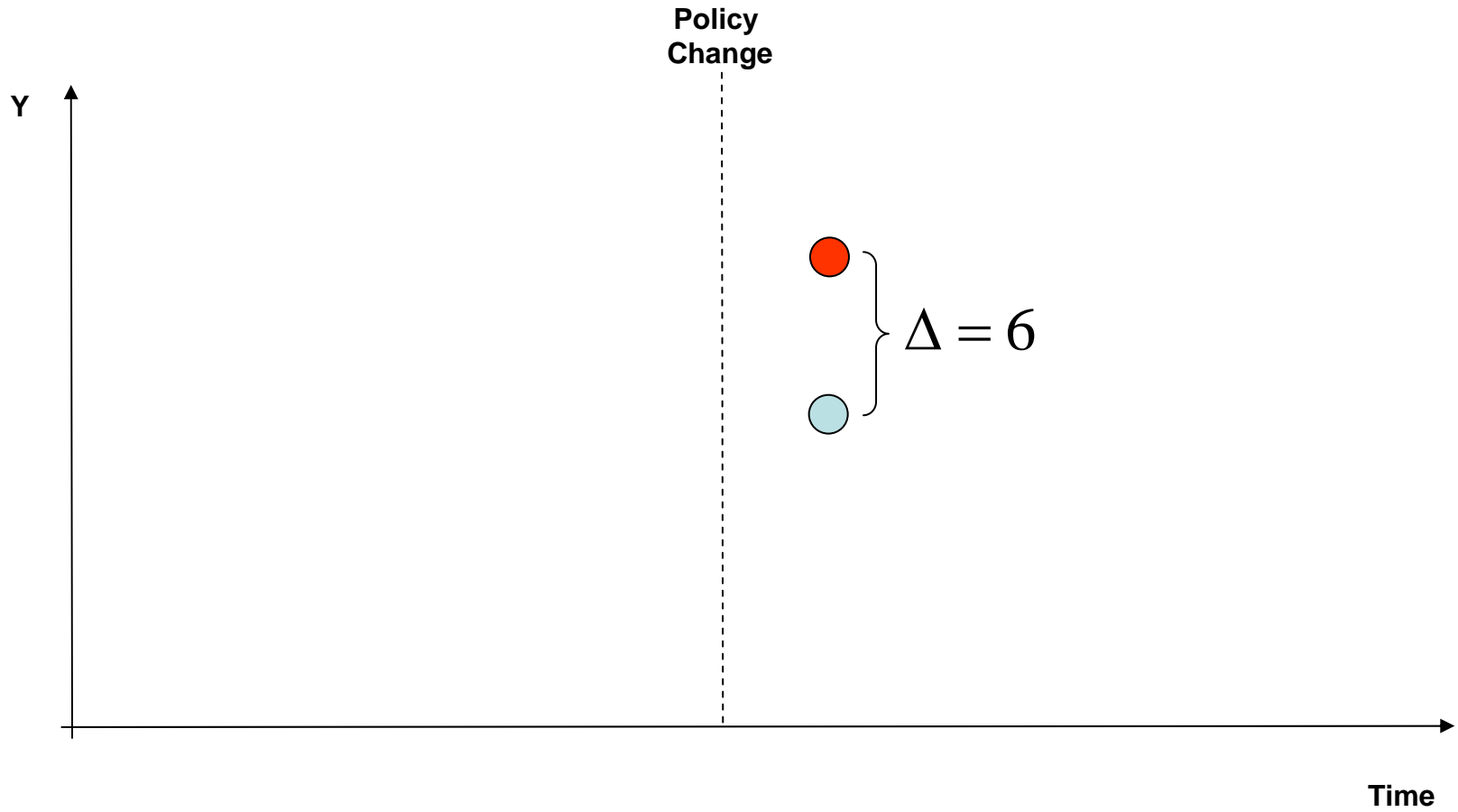
Observational Studies

If you analyze data from an observational study you cannot stop at a “statistically significant” result.

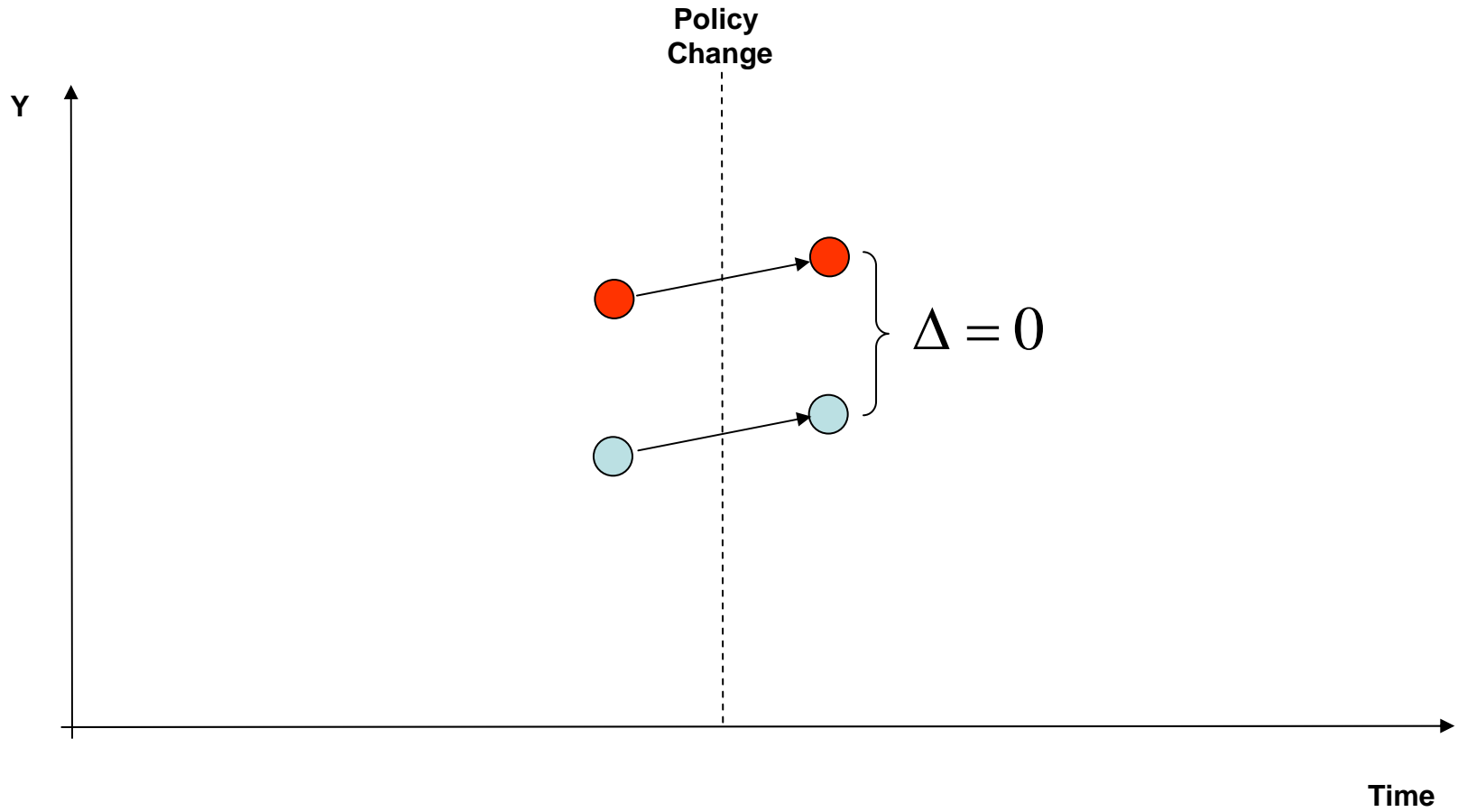
In observational studies you must work to disprove or dismiss whatever “result” you observe.

90% of the work in an observational study lies with **ruling out competing explanations. This is the key.**

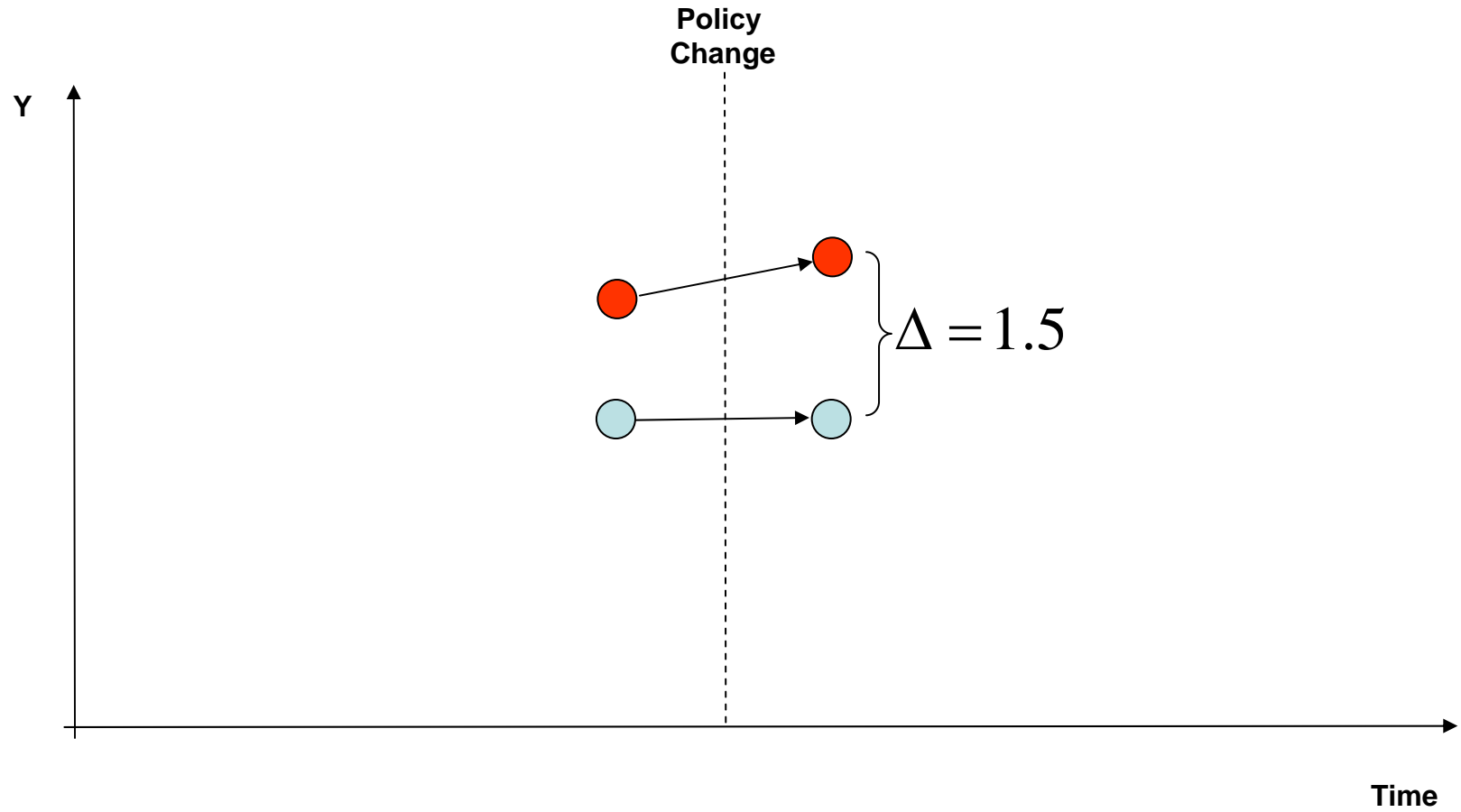
Lastly, pre/post policy change



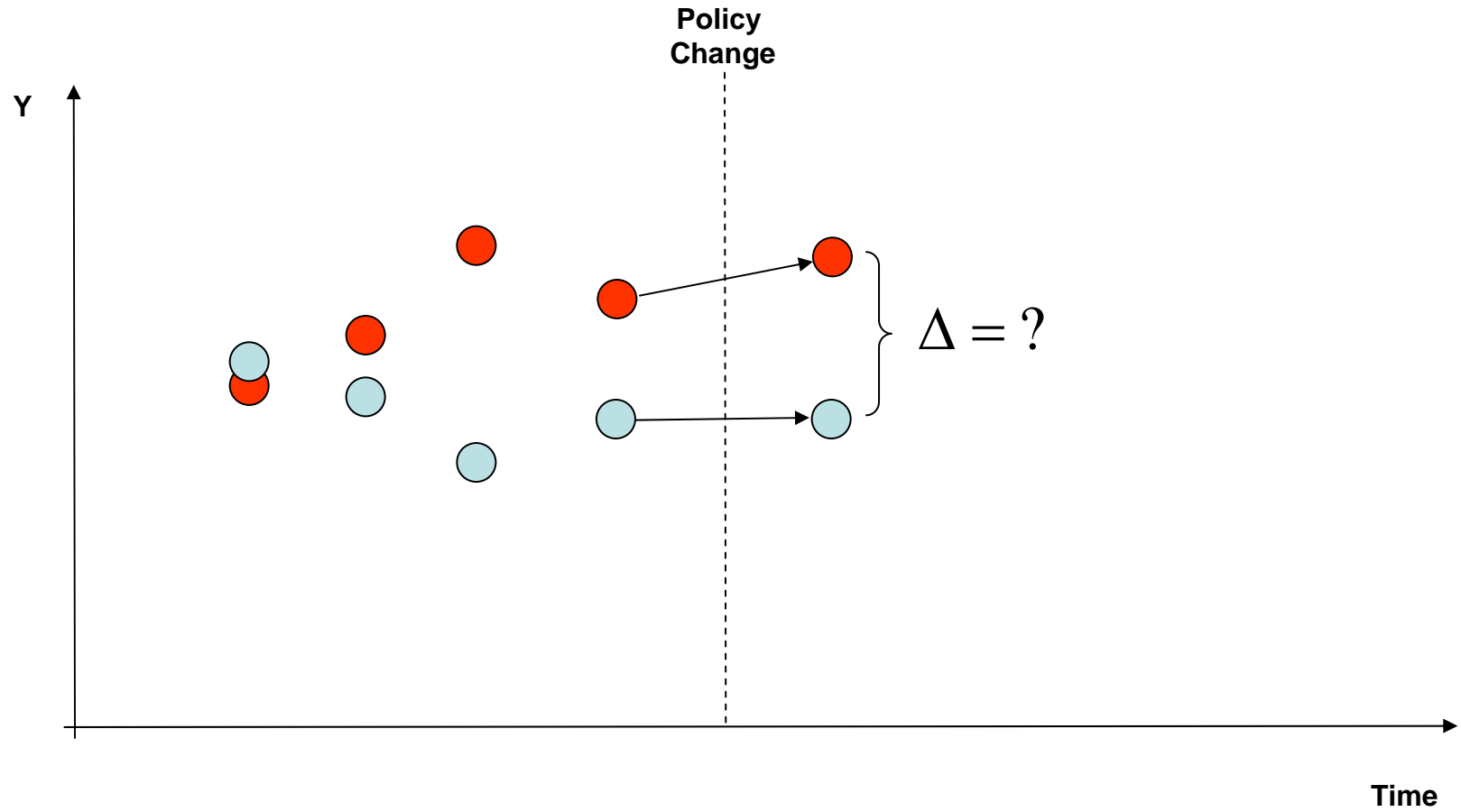
Lastly, pre/post policy change



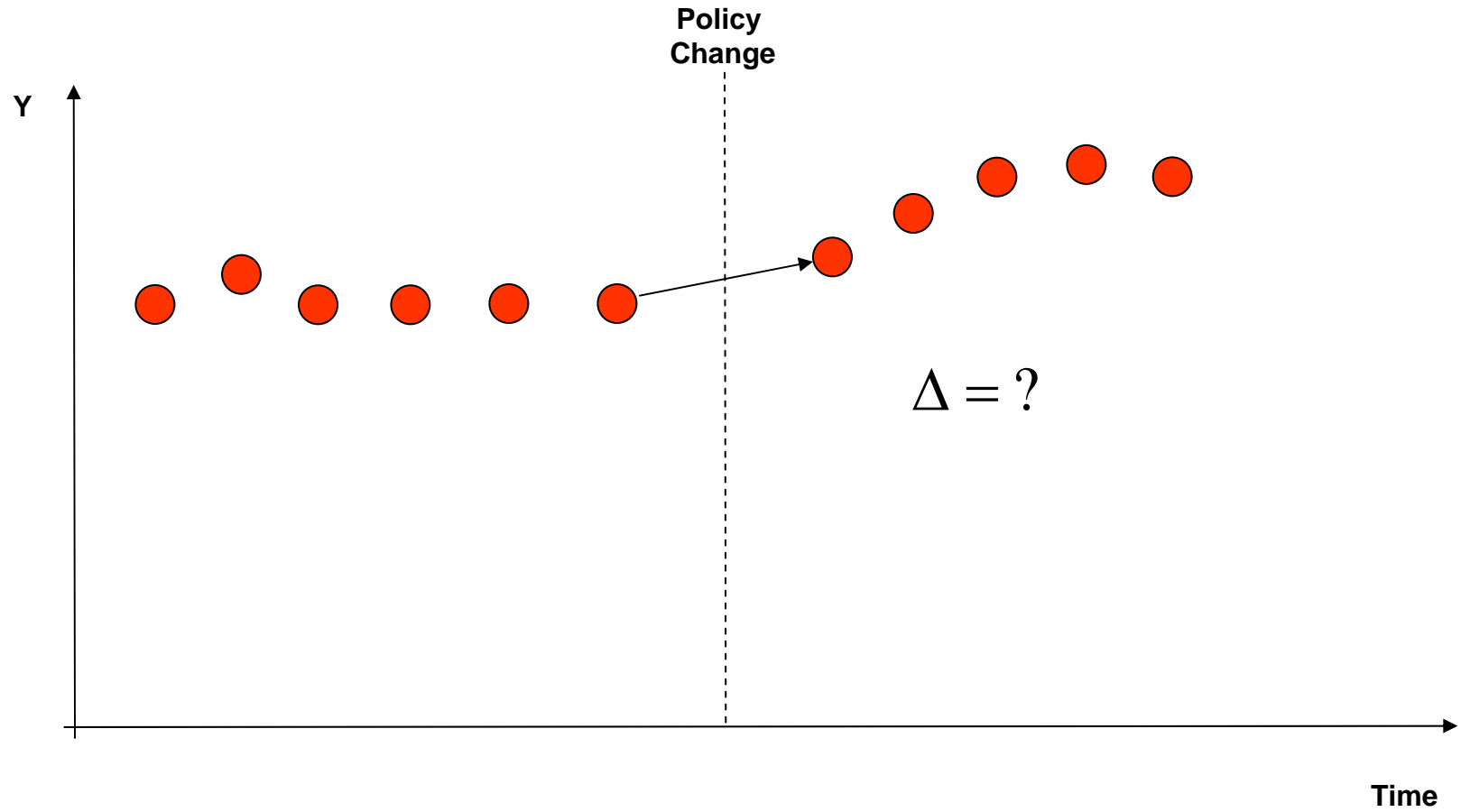
Lastly, pre/post policy change



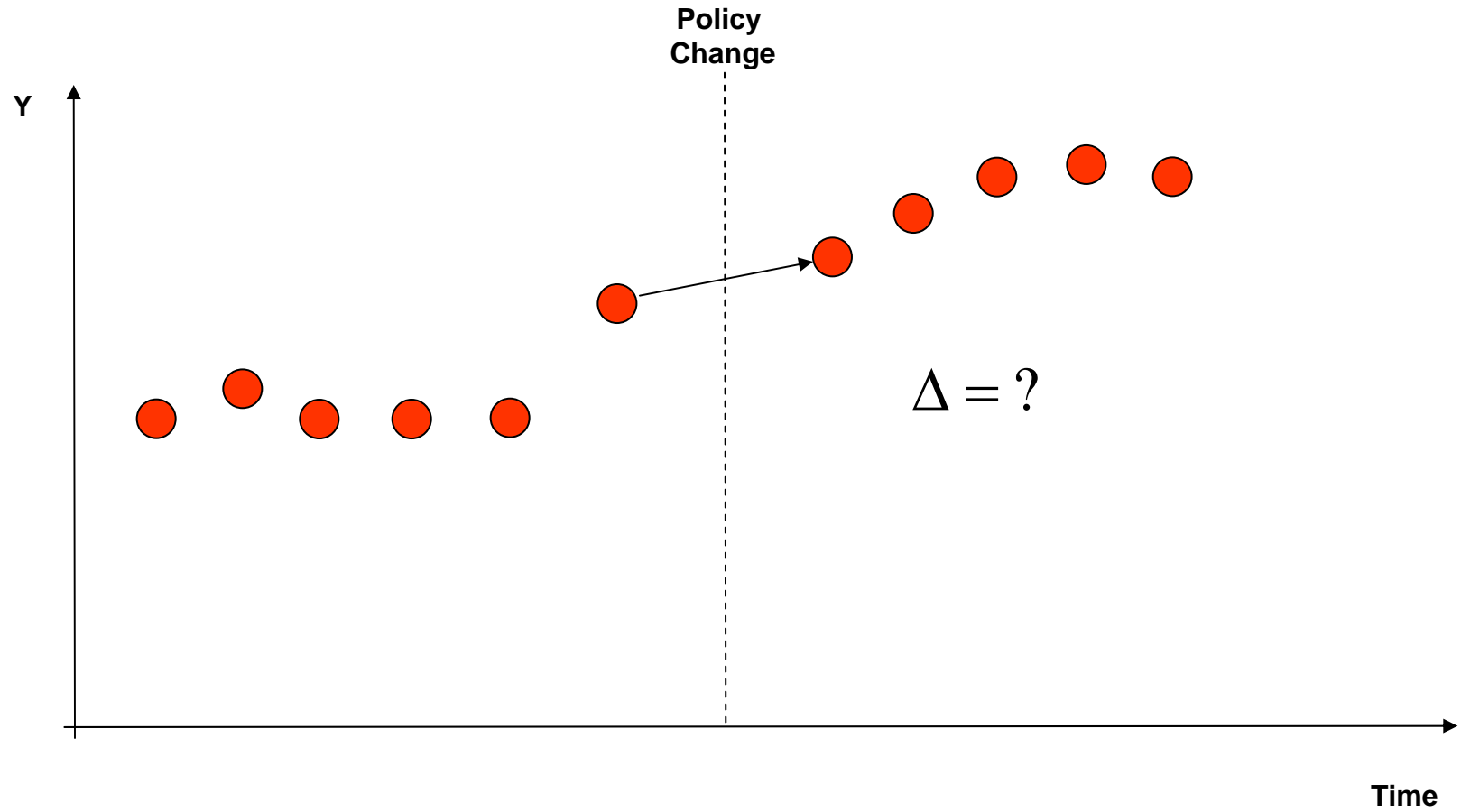
Lastly, pre/post policy change



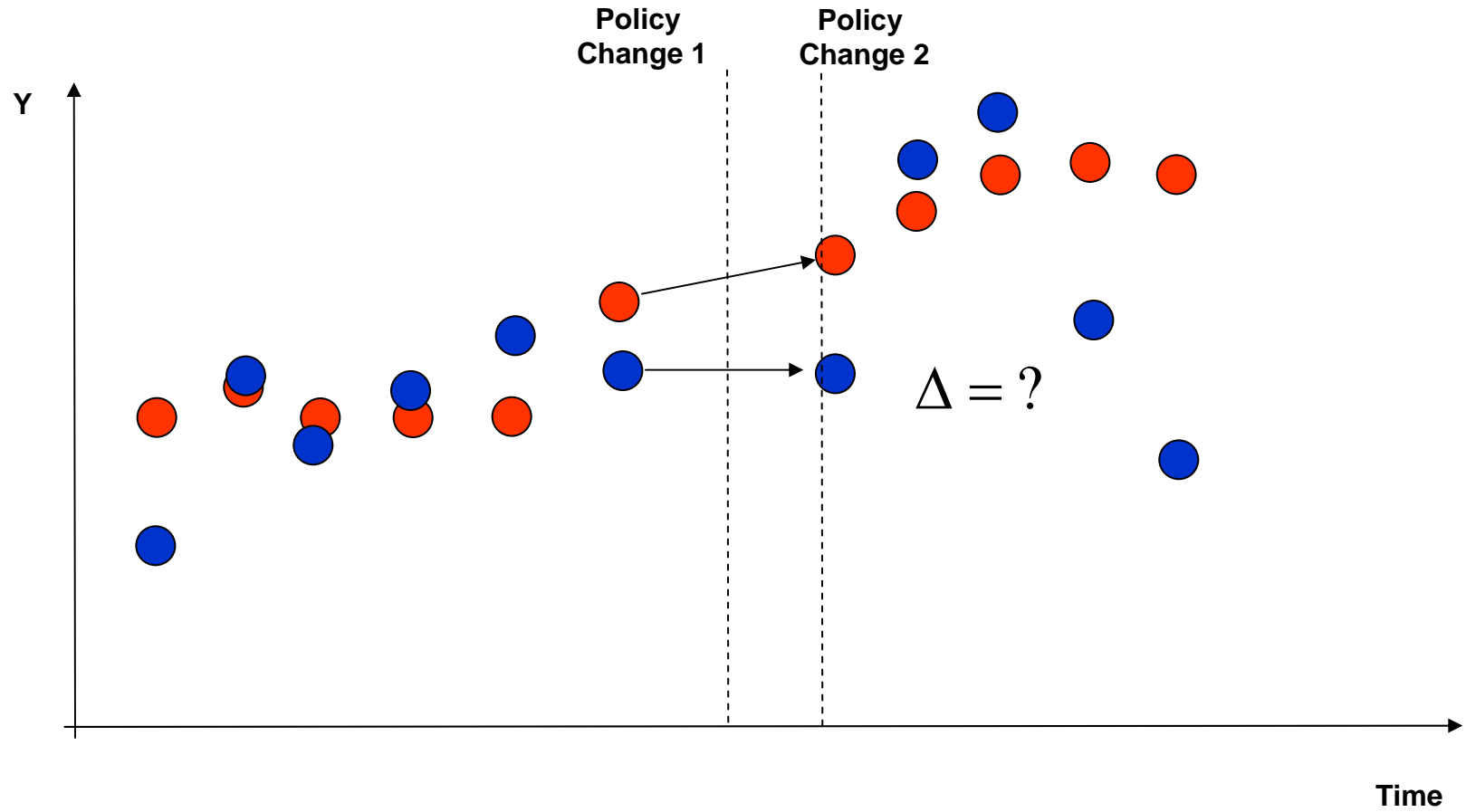
Lastly, pre/post policy change



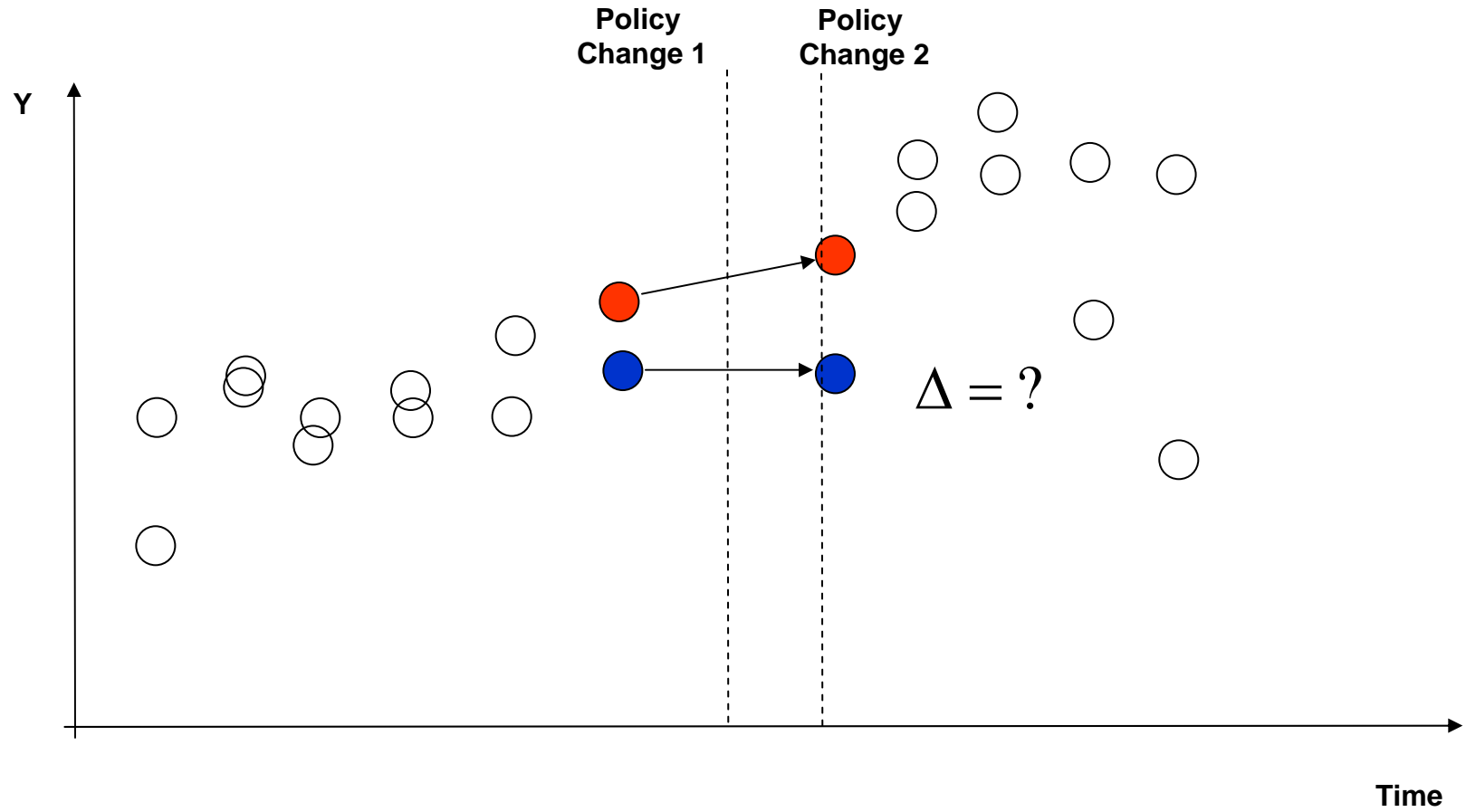
Lastly, pre/post policy change



Lastly, pre/post policy change



Lastly, pre/post policy change



In closing...

- Models substitute assumptions for data
- Seek simple designs with intense content
- Data collection matters... most?
- *Homo sapiens* are social. We must deal with it.
- To improve health: (quasi)experiments *with groups*
- Observational studies have experimental analogs
- Gell-Mann and simple stuff... like quantum physics