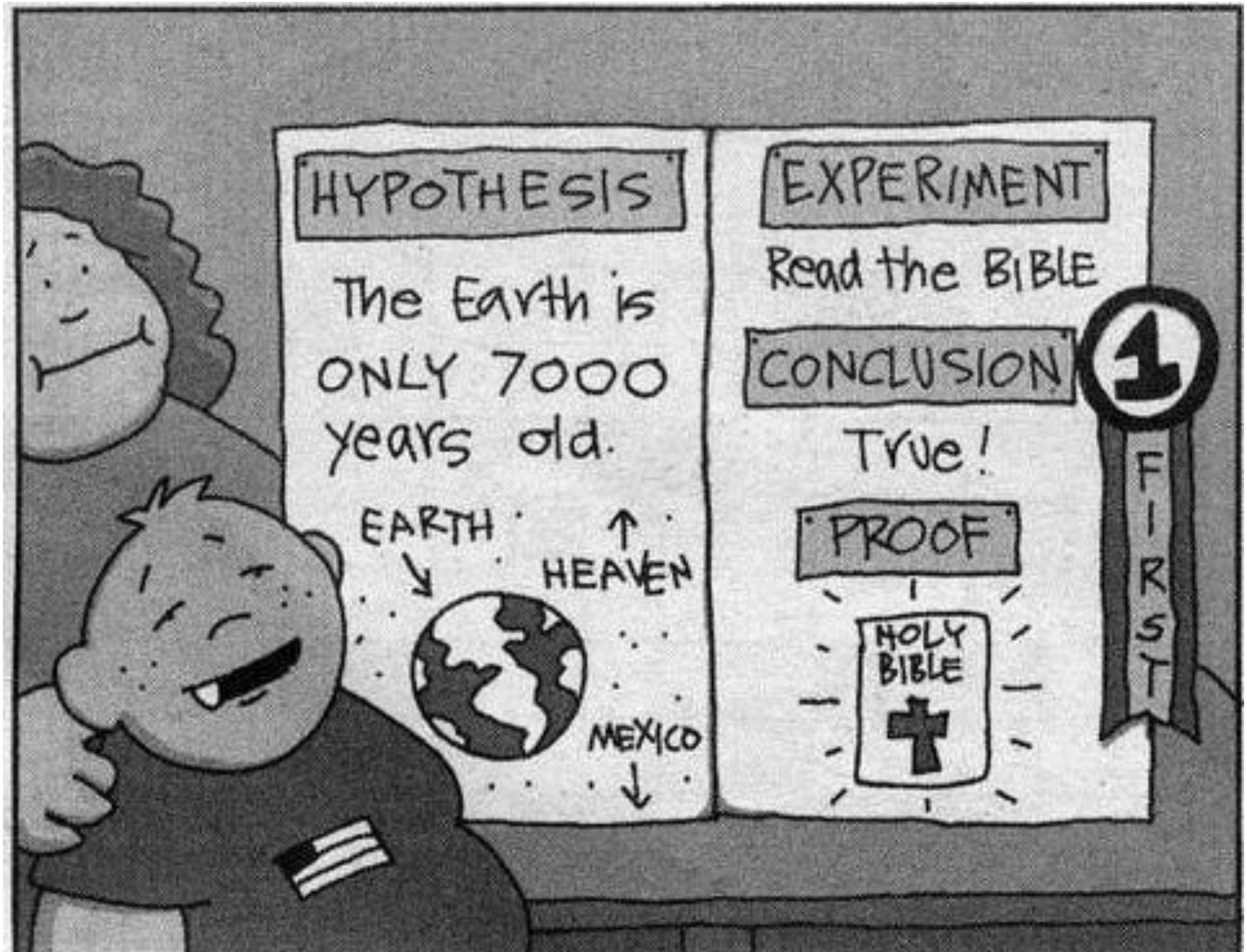


Creationist Science Fair

What is the question?

- **Effects of causes**?
  - Identifying and quantifying the causal impact of treatment(s)
    - We know it's not 0 (don't we?)
      - What size of effect do we care about?
  - What is a "treatment"?
    - No causation without manipulation?
      - What about sex, ethnicity, the Rocky Mountains…

What is the question?

- **Causes of effects**
  - We observe a regularity or pattern and ask: what are the causes of that?
  - What are the causally relevant variables and what are just the (possibly confounding) background conditions?
  - How far back and how far forward in the causal chain?
    - At what point is intervention possible?
- **Example**: Women who live close to their parents are more fertile than those who live further away. Why is that?

TABLE 3 Criteria supportive of causal inference regarding demographic change

1.	Time order	The cause should precede the effect; or, where a process is cumulative, the start of the cause should precede the start of the effect. Simultaneous causation is possible in relation to mechanical processes.
2.	Contiguity	The shorter the time between the cause and the effect, the stronger the basis for causal inference. Lags are possible, and may even be necessary, but they must be explained.
3.	Duration	Causal inference is strengthened where the effect continues during the entire period in which the cause is operating. Not always applicable—causes of very short duration may have longer-lasting effects, and some processes may be irreversible.
4.	Distinctiveness	Causal inference is more straightforward where both cause and effect are clearly differentiated and identifiable in a temporal context. True causes and effects may be hard to isolate from surrounding variability. Analogous to Bradford Hill's strength criterion, but distinctive effects may not be large and vice versa.
5.	Direction	The effect should be in the expected direction—i.e., would the effect and its direction have been predicted before the event? Unexpected effects may occur and expected effects may be hard to specify. Analogous to Bradford Hill's plausibility criterion.
6.	Proportionality	Causal linkage is better grounded when the scale of the effect can be considered proportional to the scale of the cause. Need not always apply—the criterion is subjective: apparently small causes can have major effects, and the reverse may also hold.
7.	Recurrence	Causal inference is strengthened if the linkage occurs in a variety of settings. Context may, however, preclude exact replication. Not essential—some causes are historically unique. Analogous to Bradford Hill's consistency criterion.
8.	No cause, no effect	Where the putative cause is absent, the effect is absent too. May not always apply in that multiple causes of a given event are possible.
9.	Mechanism	To establish a causal link, a plausible set of intermediate links is required showing how the cause brings about the effect. Specifying and providing evidence of the mechanism involved is essential.
10.	No alternative	All reasonable alternative explanations, including confounding, must be considered and ruled out. This criterion is simpler to satisfy where effects are large and distinctive. What is considered a reasonable alternative may change through time.

What is the question?

- **Something else?**
- Population heterogeneity?
  - Q. Is A “the same” as B? A. No (We know without looking). Tell me when “not the same” is great enough to matter.
    - Social mobility
    - Ethnic differences in educational attainment
  - In what ways do groups differ from one another?
  - Is something the case or is something else the case?
- A lot of empirical sociology has this flavour

Instrument: a source of exogenous variation in T

$$\begin{tikzcd} R \arrow[dr] & & \\ & T \arrow[r] & Y \end{tikzcd}$$

Classic randomized experiment

$$\begin{tikzcd} I \arrow[dr] & & \\ & T \arrow[r] & Y \end{tikzcd}$$

“Natural” experiment – draft lottery

$$\begin{tikzcd} I \arrow[dr] & & \\ & T \arrow[r] & Y \end{tikzcd}$$

Month of birth as instrument for years of education in estimating returns to education

$$\beta_{ols} = Cov(TY)/Var(T)$$

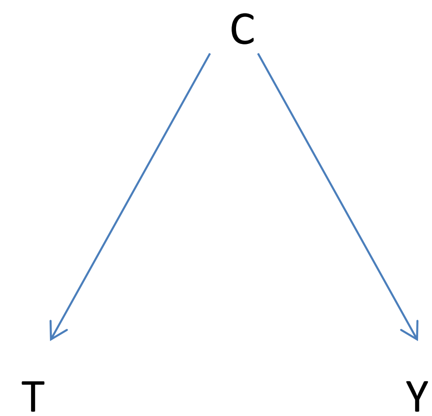
$$\beta_{IV} = \beta_{IY}/\beta_{IT}$$

No instrument? Then model selection into treatment

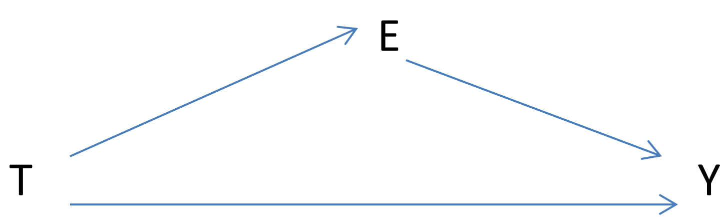
- Two Approaches
  - Condition on all the observable factors that control allocation to treatments
  - Match (balance) treatments with respect to confounders
    - Ex ante blocking
    - Ex post matching

Condition on all the observable factors

Remember: the point is to estimate the causal effect of T on Y



Usually it will not be a good idea to condition on another outcome of T ie an endogenous variable. You won't estimate the causal effect of T on Y and you will possibly introduce some backdoor selection.



NB There is a big literature on so called "mediation" analysis. It is well intentioned but often misguided

Match on all the observable factors

- Old tradition and lots of new variants
- Data demanding
- Are there enough good matches?
- How close is close enough?
- Draws attention to region of common support
  - ie must have cases that are similar on the Xs **and** have observed values for **both** treatment and control
    - Make more limited claims
- Also possible to combine matching and conditioning (regression)

Match on all the observables

- Remarkably it turns out to be sufficient to match on a function of the Xs – the propensity score
- Predict who gets the treatment with a logistic regression
- Probability of being in the treatment group has all the relevant information.
- Match on that
- NB This does not “control” for selection on unobservables