Confounding: Casual versus Causal Analysis

David A. Kenny

University of Connecticut

http://w3.nai.net/~dakenny/kenny.htm

Outline

- Meaning of bias in this talk
- General strategies for reducing bias
- The selection of variables for statistical control
- Effects of statistical control
- Strange cases

What Types of Bias Are NOT Discussed

- Sampling Bias
- Statistical Bias
- Experimenter Bias
  - Communication to participants
  - Data analysis choices
- Bias in the "Signal Detection" Sense: Tradeoffs

What is Bias (in this talk)?

- Estimated Effect = True Effect + Error
- Error = Systematic + Random
- Systematic Error is BIAS
- Very often called “Confounding”
- Sometimes called “Selection” or “Omitted Variable Problem”

How to Control for Bias

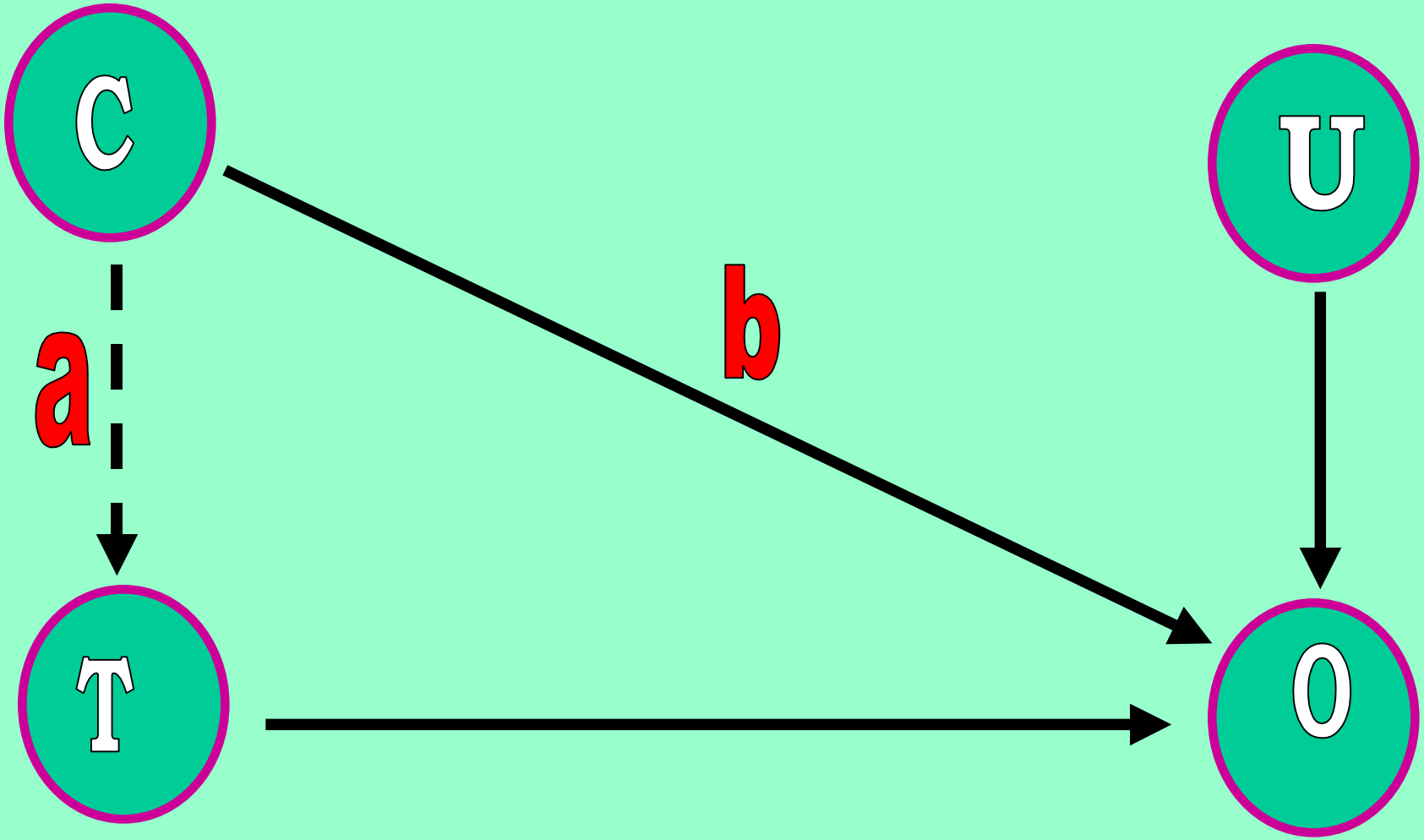
- Design
  - Random assignment
  - Hold constant
  - Stratification
- Statistical Analysis

Avoid doing something statistically that could be done by design.

Notation

- **T: Treatment**
- **O: Outcome**
- **C: Control Variable**
- **U: Other Causes of O besides T and C (not always drawn)**
- **X: Hidden Variable or a Variable that Should be Controlled but is Unmeasured**

Model of Statistical Control



Strategies for the Selection of Confounding Variables

- Lore or the usual suspects (but be critical)
- Outcome-oriented selection
- Treatment-oriented selection: Propensity Scores
- Do not control for a mediator

Effects of Statistical Control in Linear Regression

- **Adjustment**
  - **new effect = old effect - ab**
- **Precision**
  - **increased if error variance is reduced (b nonzero)**
  - **reduced if the control variable is correlated with the treatment (a nonzero)**
  - **reduced due to a loss of a degree of freedom**

Effects of Statistical Control in Randomized Studies

- Case 1 (stratification): **a** is zero
  - likely increase in precision if **b** is nontrivial in size; there is no adjustment
- Case 2: **a** is close to zero, but not zero
  - likely increase in precision if **b** is nontrivial in size; there is a small adjustment
- Case 3: **a** is clearly non-zero
  - uncertain effect on precision , depends on **b**; there is a large adjustment

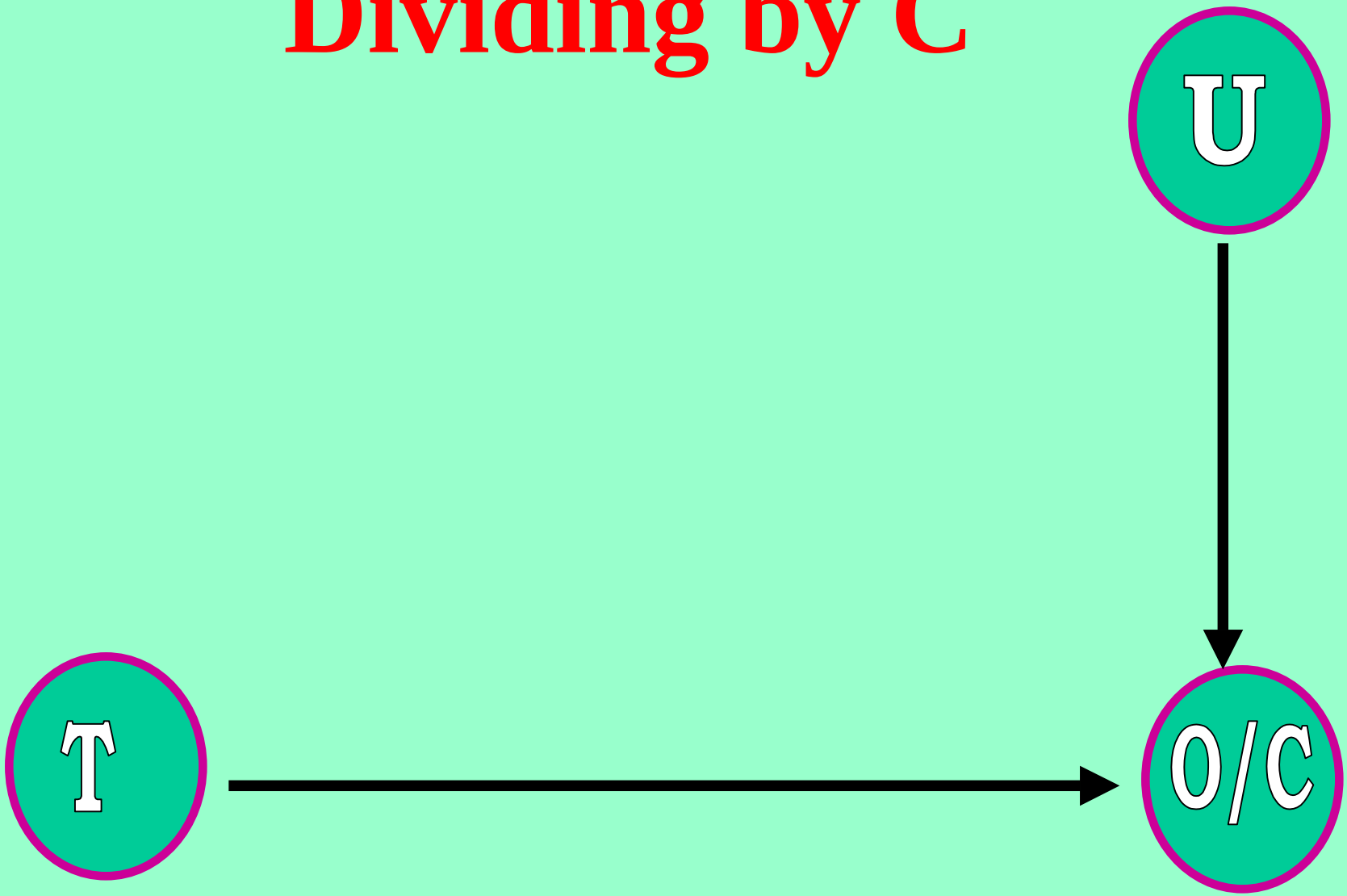
Effects of Statistical Control in Non-randomized Studies

- a nonzero and b small
  - loss of precision
  - little adjustment
- ab nonzero
  - precision gain or loss depending on the relative size of the a and b
  - adjustment which is likely helpful even if precision is reduced

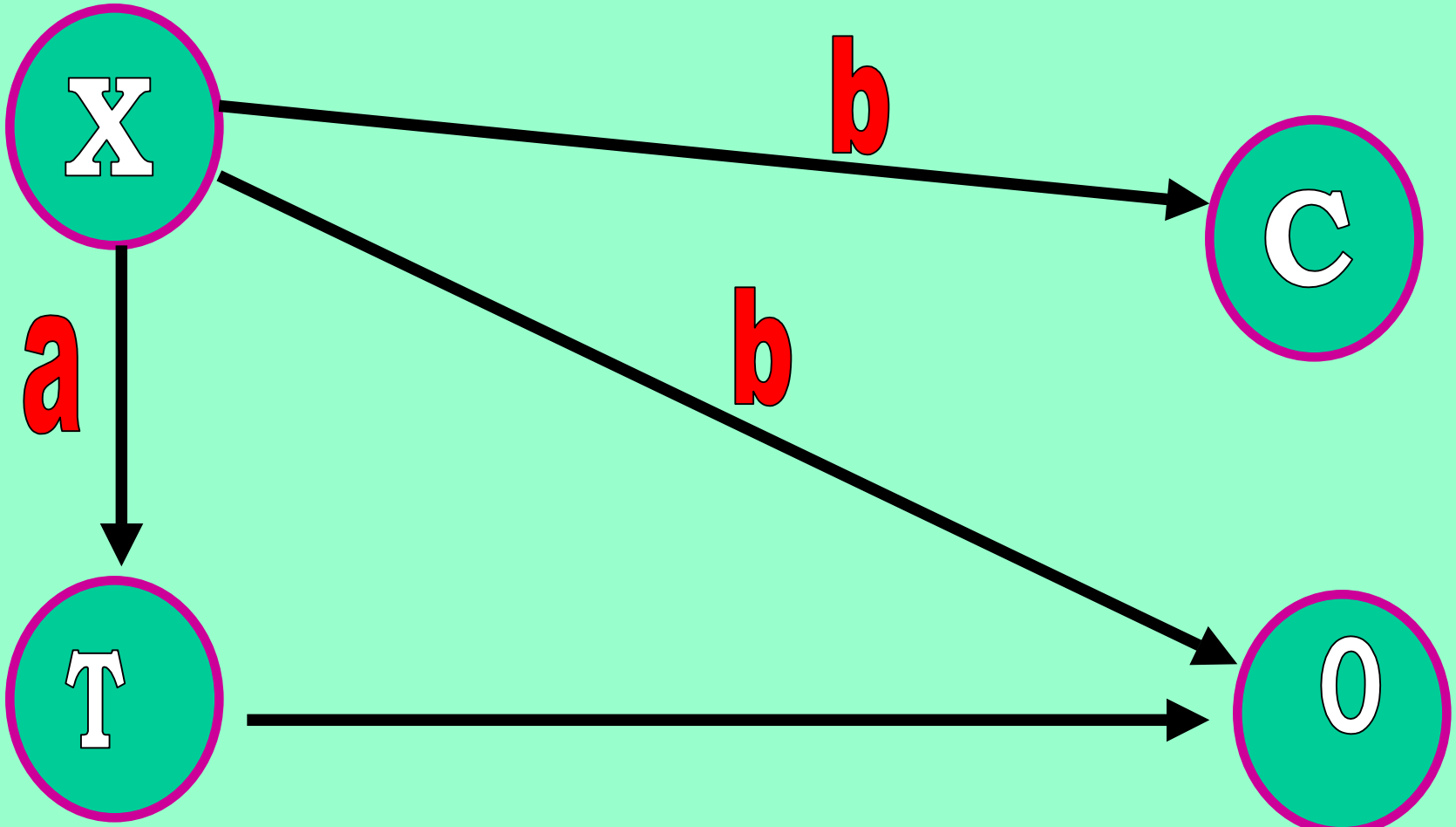
Key Decisions in Statistical Control

- functional form
  - linear
  - multiplicative
- choice of estimation strategy
  - statistical estimation (put on the right side)
  - a priori (put on the left side)
    - base-rate correction
    - common confound
- measurement error

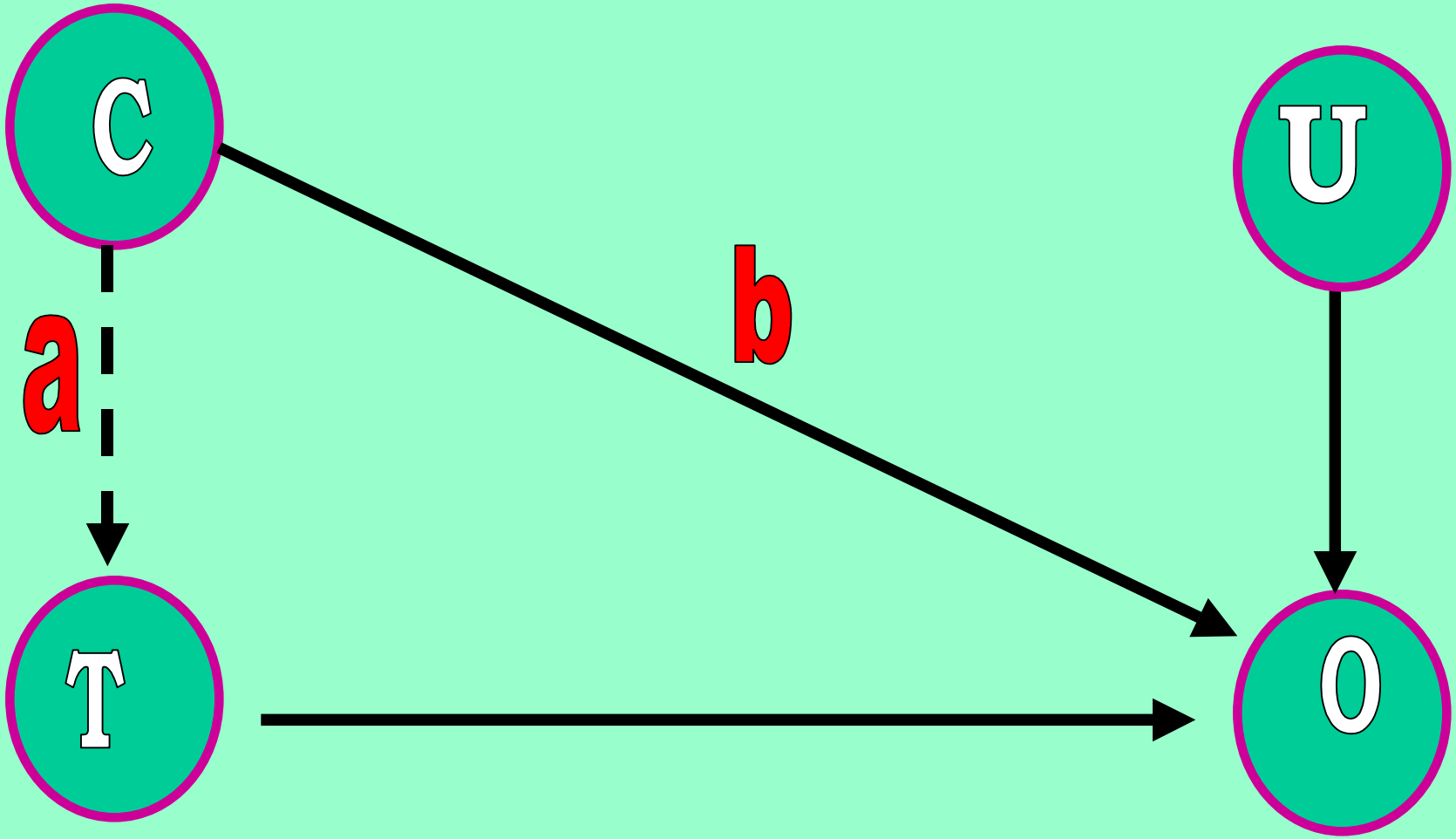
Base-rate (C) Correction: Dividing by C



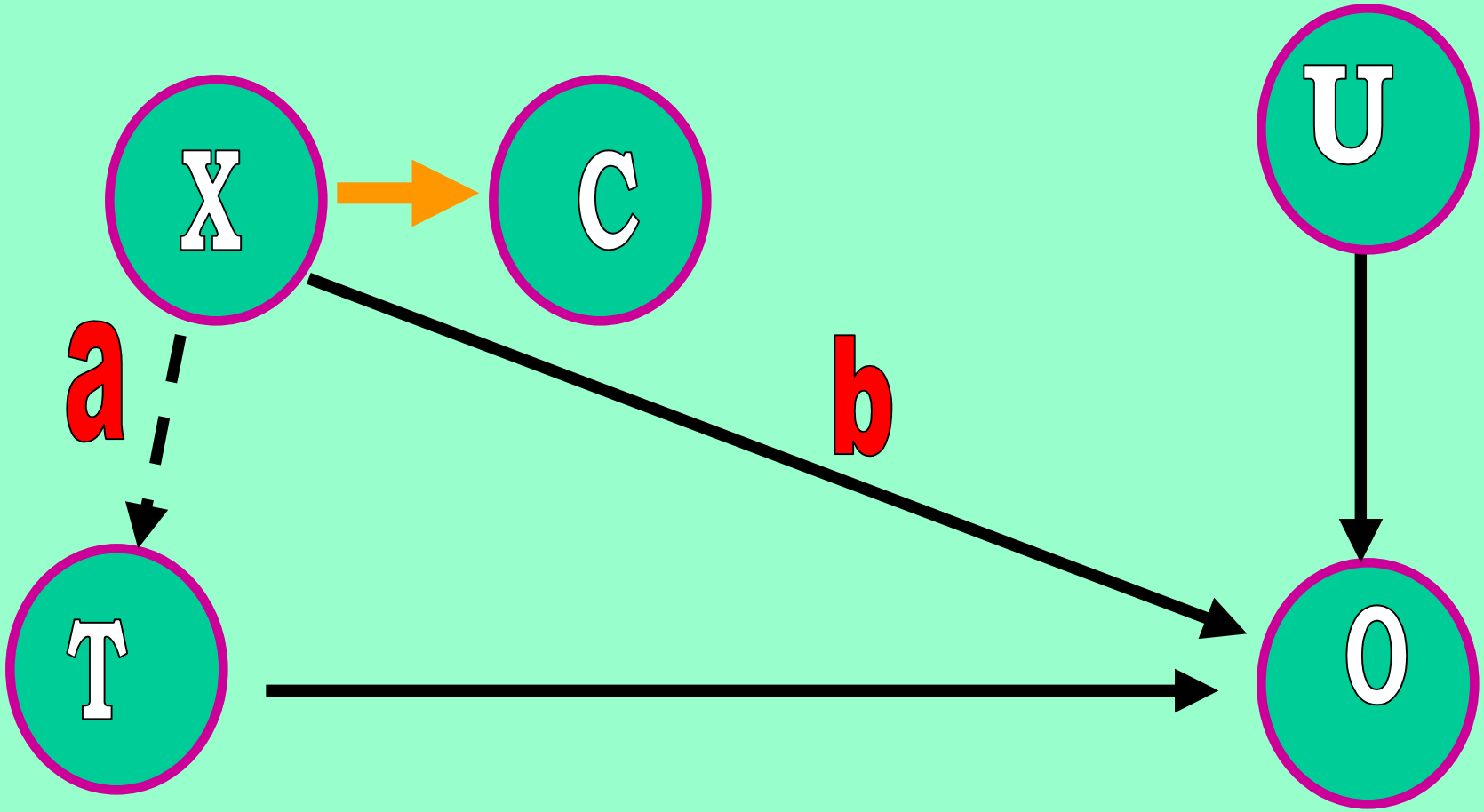
Model for the Use of Difference Score (O - C) for Statistical Control



Model of Statistical Control



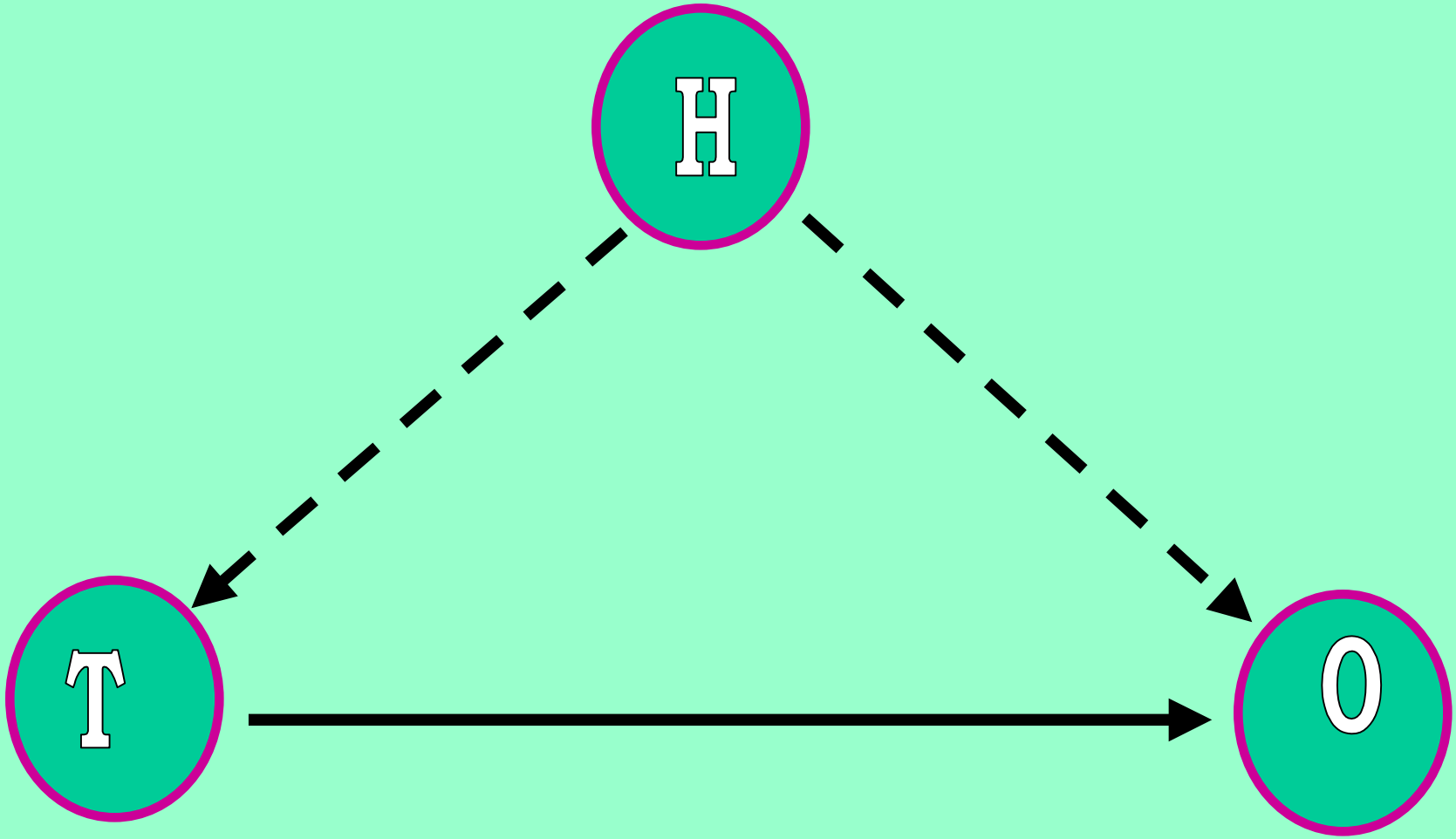
Measurement Error in a Control Variable (C)



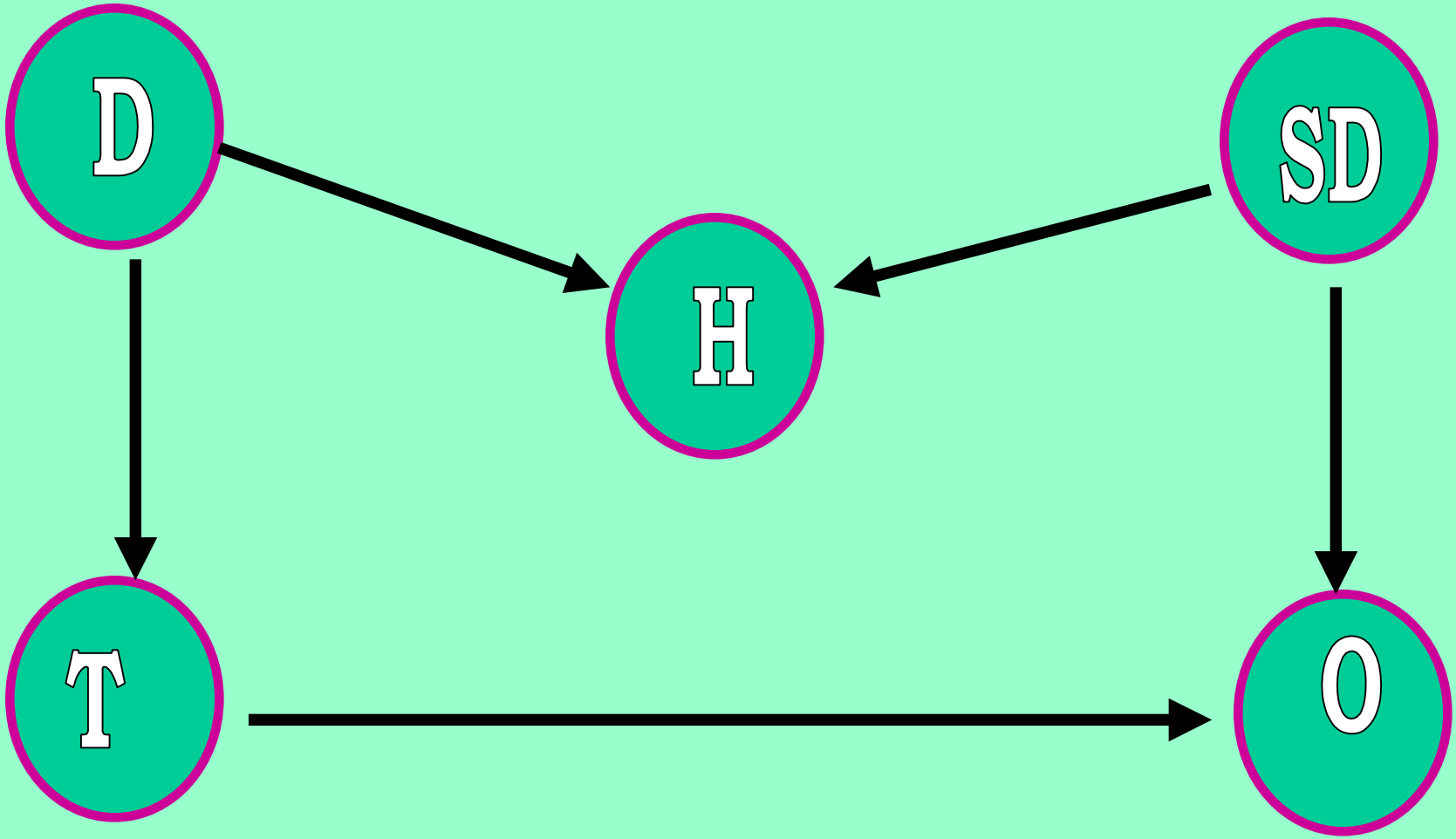
Strange Cases: Bias Created by Statistical Control

- **Case I: Correlation is not Causality: Just Because a Variable (H) Correlates Does Not Necessarily Make It a Good Control Variable**
- **Case II: Two Wrongs Make a Right or Doing Good Leads to Bad: Control for a "Confound" May Create Bias**

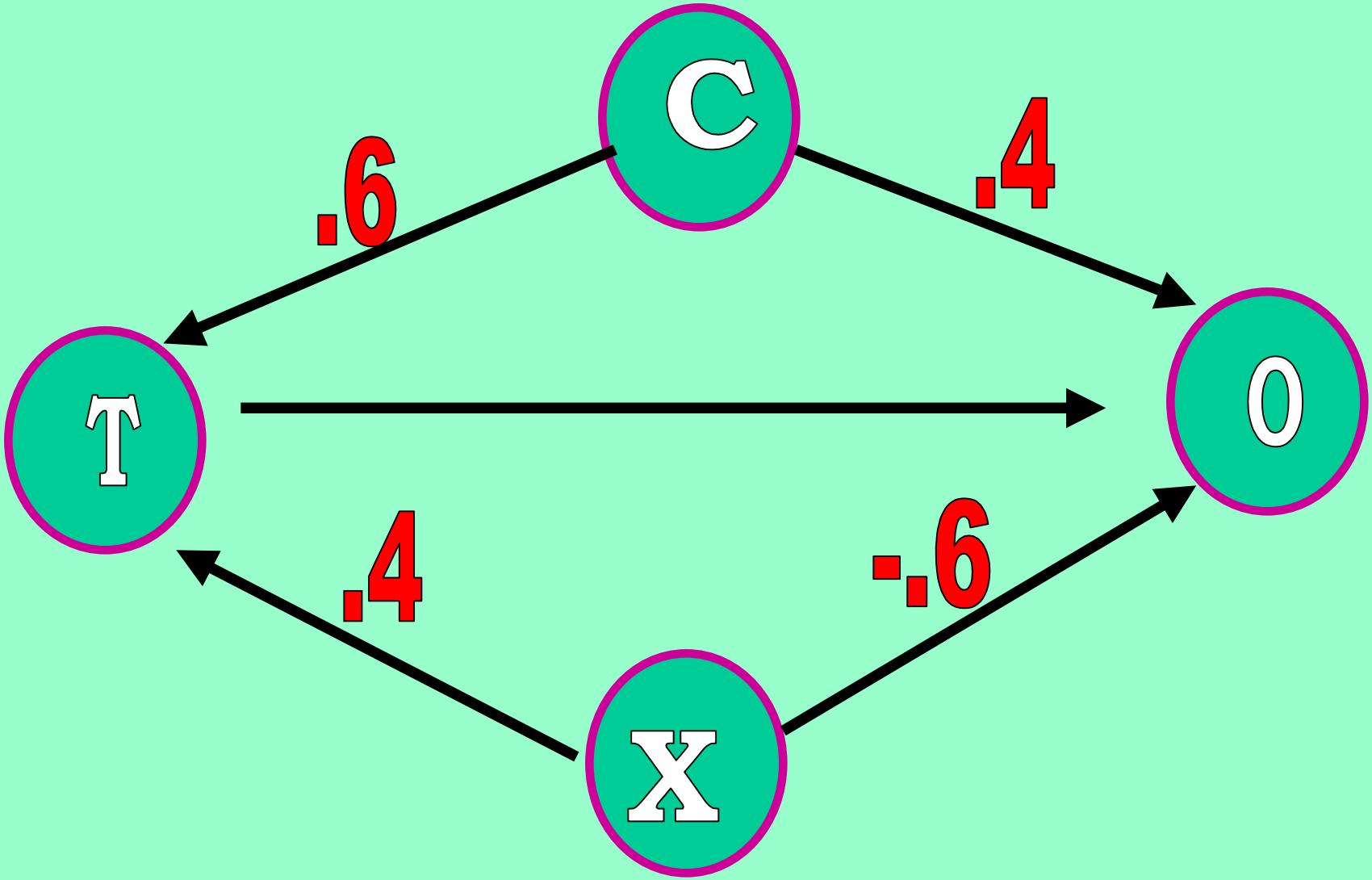
Correlation is Not Causality



Correlation is Not Causality



Two Wrongs Make a Right



Conclusions

- **Do not casually “control” for a variable by entering it into the equation.**
- **An understanding of the causal process is necessary to adjust properly for variables.**
- **Know whether your goal is to adjustment or precision.**
- **Know more about the variables of interest or know people who do.**

References

- Cook & Campbell, *Quasi-experimentation*, Chapters 1-4.
- Pearl, *Causality*, Chapter 6.
- Campbell & Kenny, *Primer on Regression Artifacts*, Chapters 5-6.
- APA Taskforce, *American Psychologist*, August 1999, pp. 594-604.