# Instrumental variables
# Observational settings

INFO/STSCI/ILRST 3900: Causal Inference

23 Oct 2025

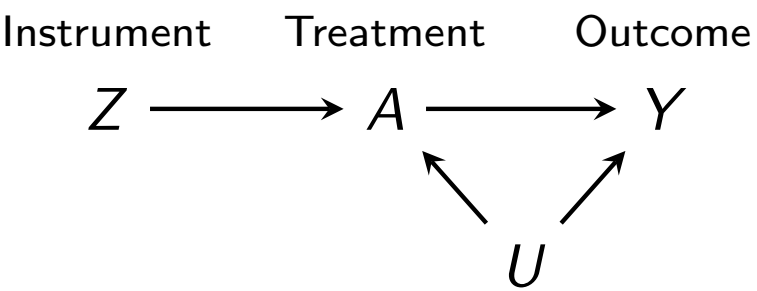
# Logistics

- Pset 4 due Oct 27
- Project Groups will be assigned

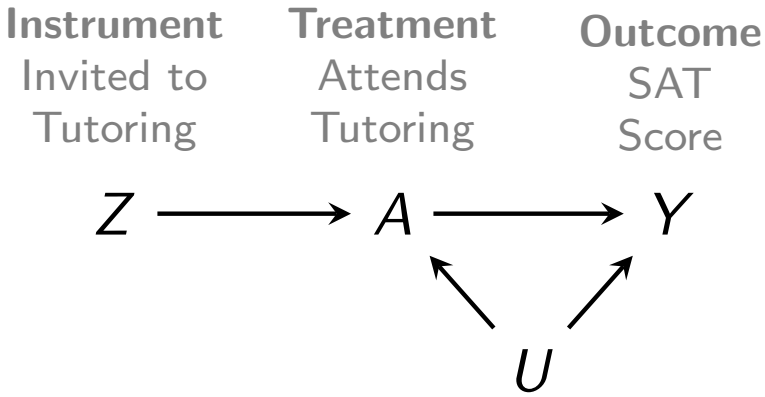
# Learning goals for today

At the end of class, you will be able to:

1. Explain requirements for IV in observational settings
   - instrument exchangeability: $Z \to Y$ is identified
   - no direct effect: $Z \to Y$ is entirely $Z \to A \to Y$
   - relevance: $Z \to A$ is sufficiently large
   - monotonicty: instrument affects treatment in one direction
2. Argue for when these hold or do not hold



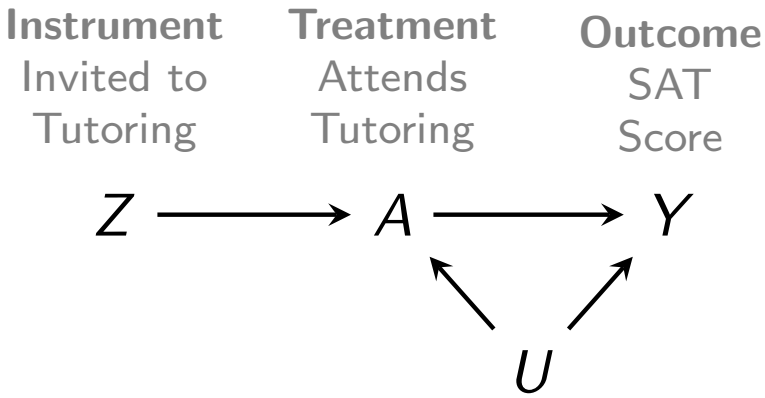
# Running example




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Running example

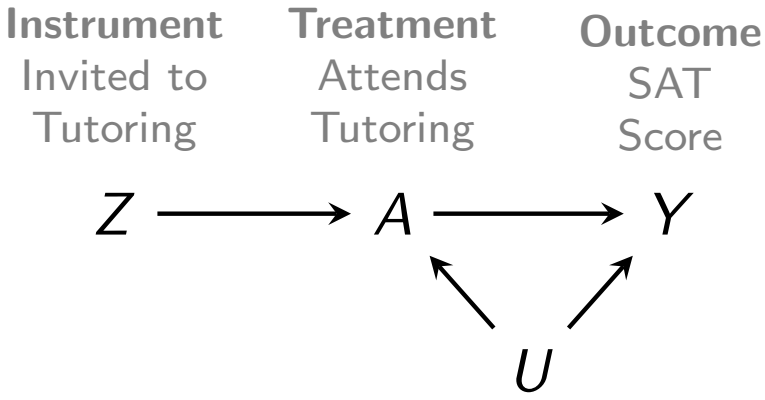



The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

**This procedure relies on the DAG being correct!**

# Running example



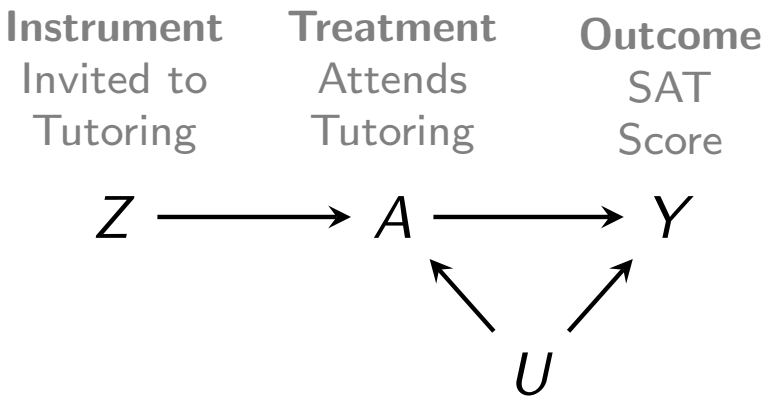

The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

**This procedure relies on the DAG being correct!**

Discuss: Are there other edges that might be in the graph?

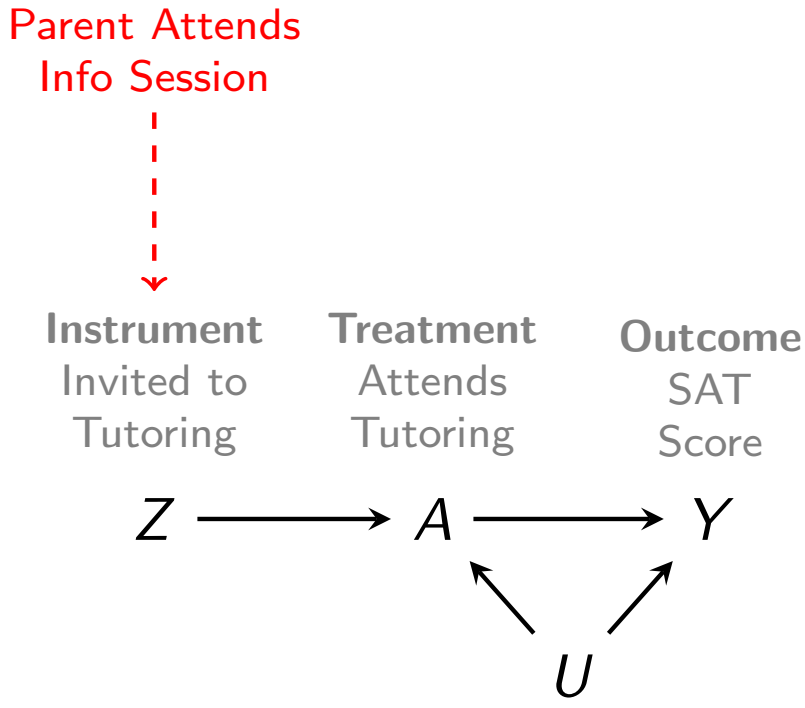
# Requirement 1) Instrument exchangeability




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

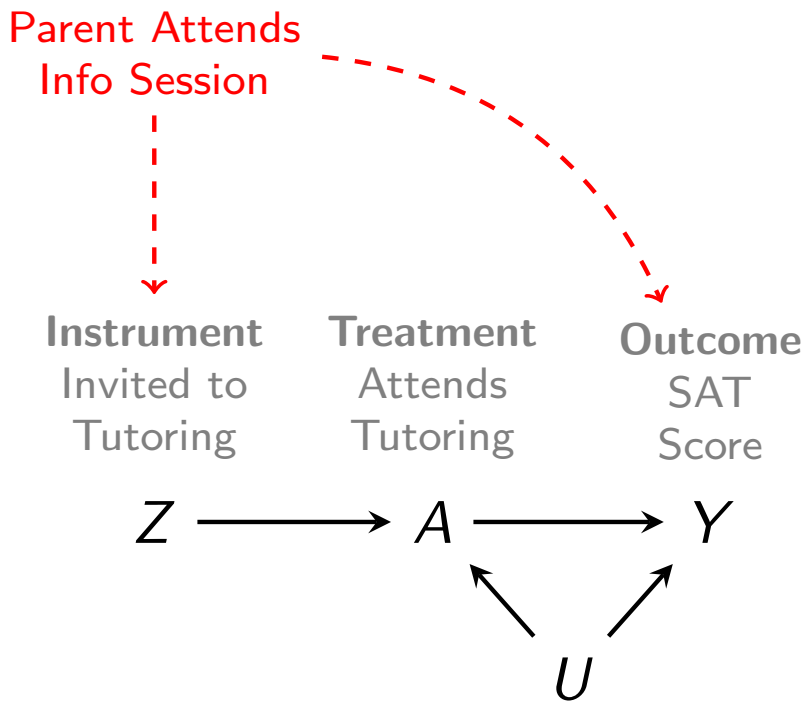
# Requirement 1) Instrument exchangeability




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

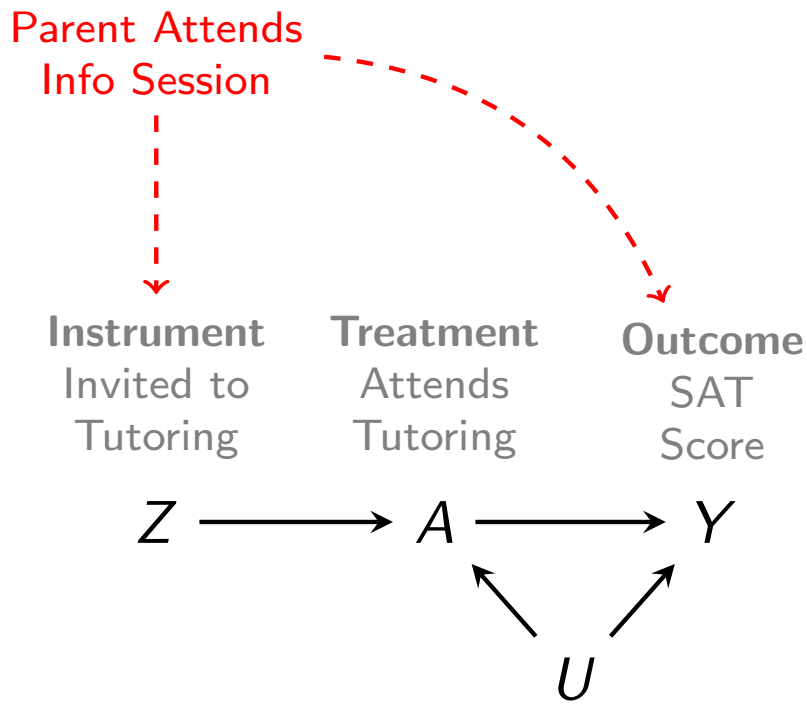
# Requirement 1) Instrument exchangeability




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Requirement 1) Instrument exchangeability

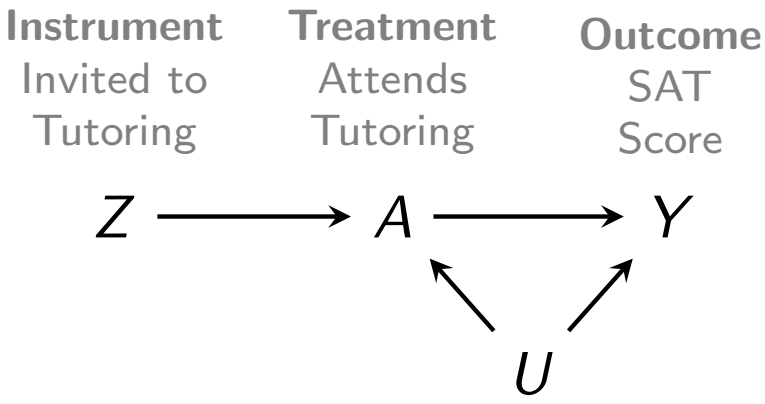



The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

doesn't work!

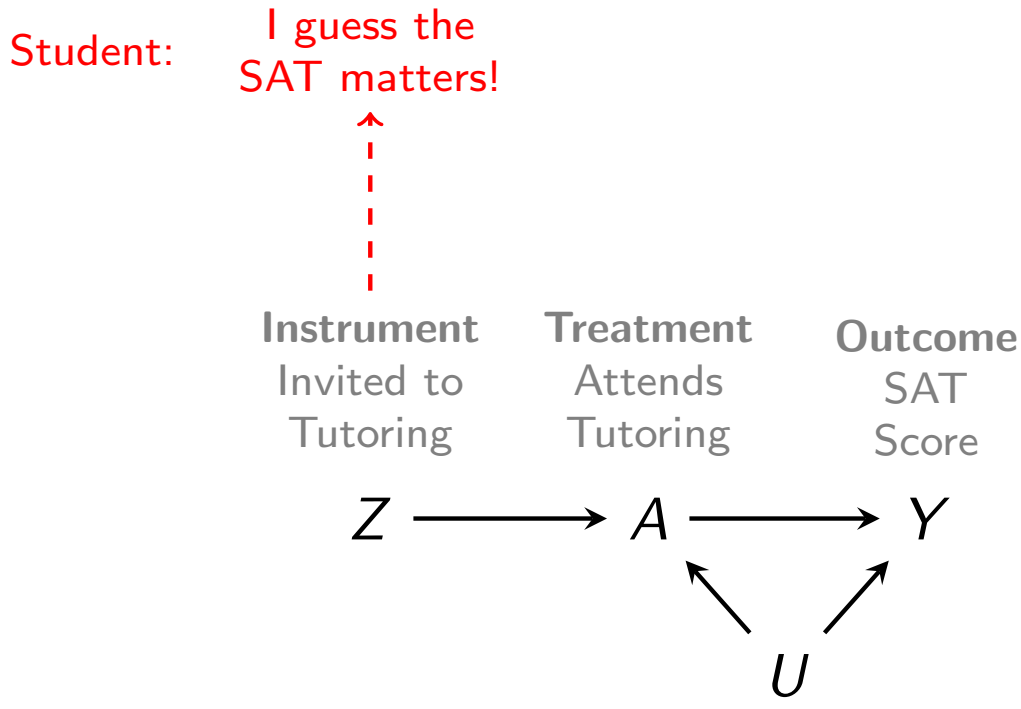
# Requirement 2) No direct effect of $Z$ on $Y$




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

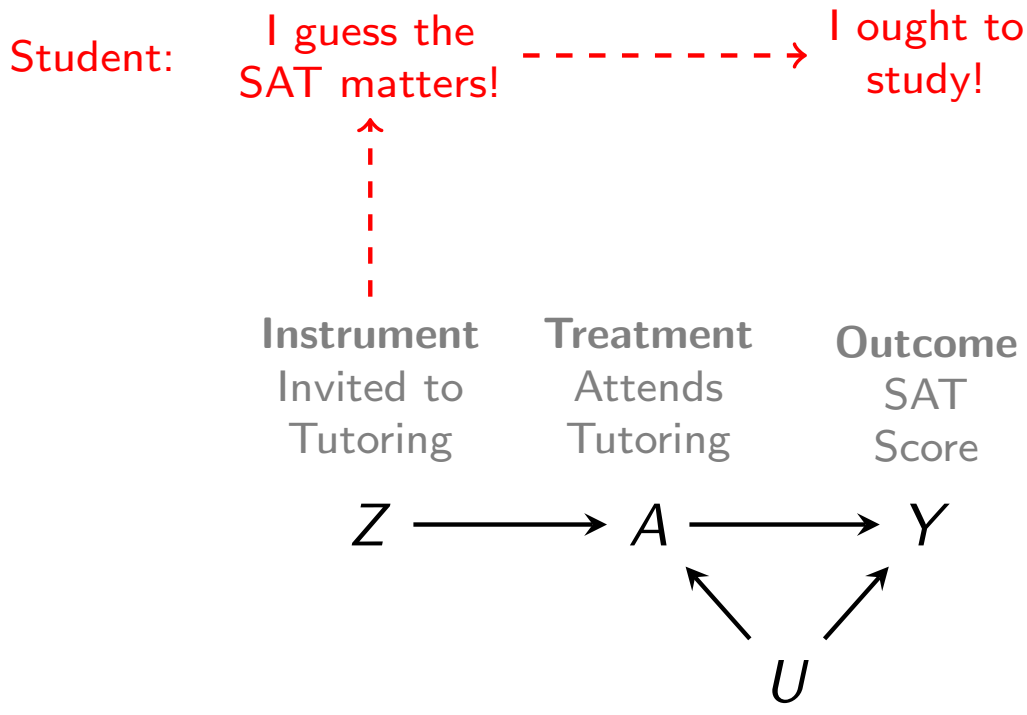
# Requirement 2) No direct effect of $Z$ on $Y$




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

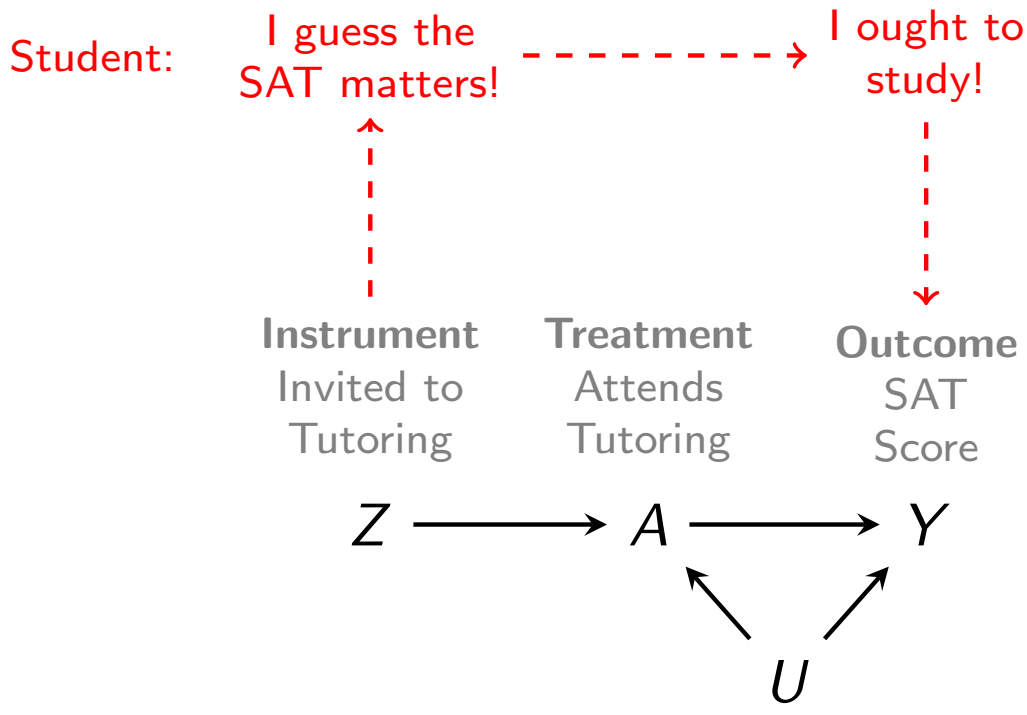
# Requirement 2) No direct effect of $Z$ on $Y$




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

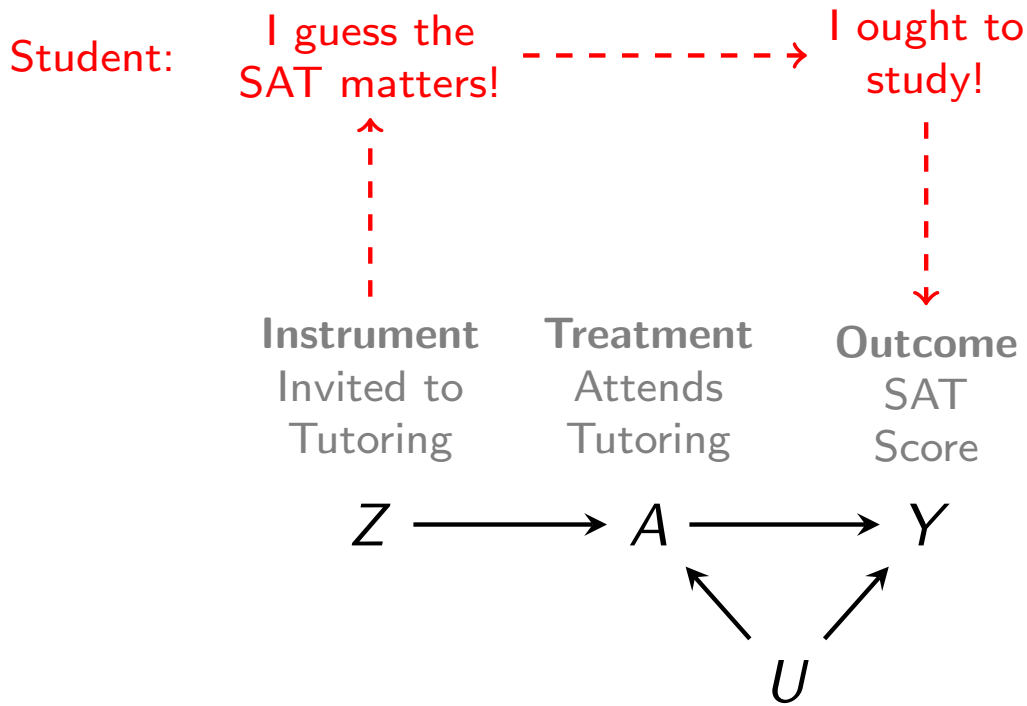
# Requirement 2) No direct effect of $Z$ on $Y$




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Requirement 2) No direct effect of $Z$ on $Y$

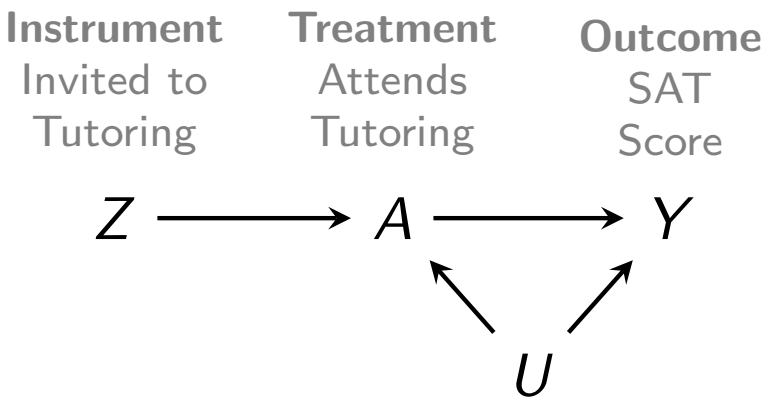



The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

doesn't isolate $A \rightarrow Y$

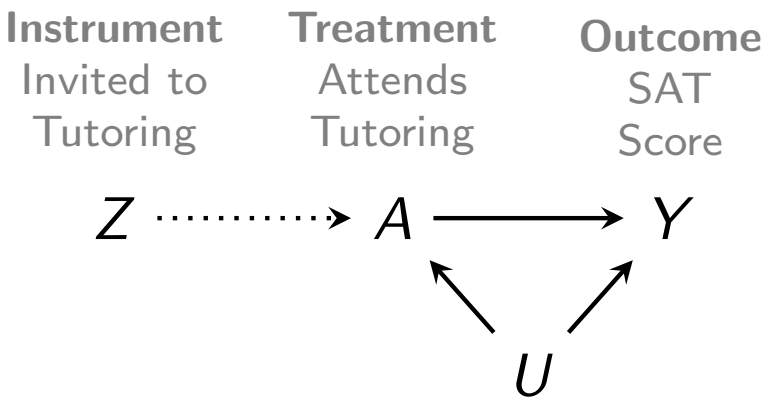
# Requirement 3) Relevance




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Requirement 3) Relevance

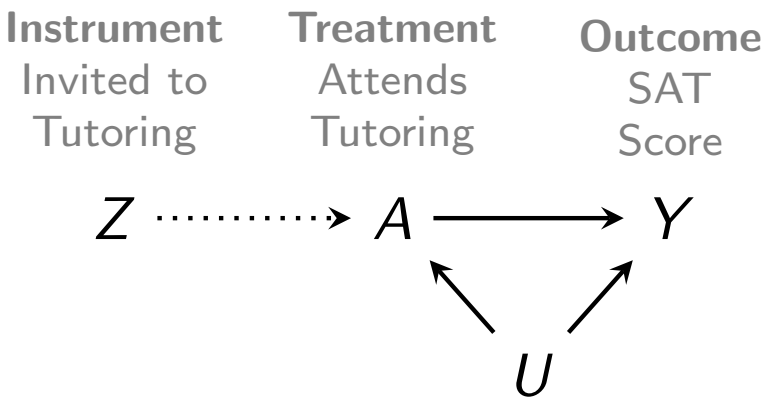



The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

hardly any!

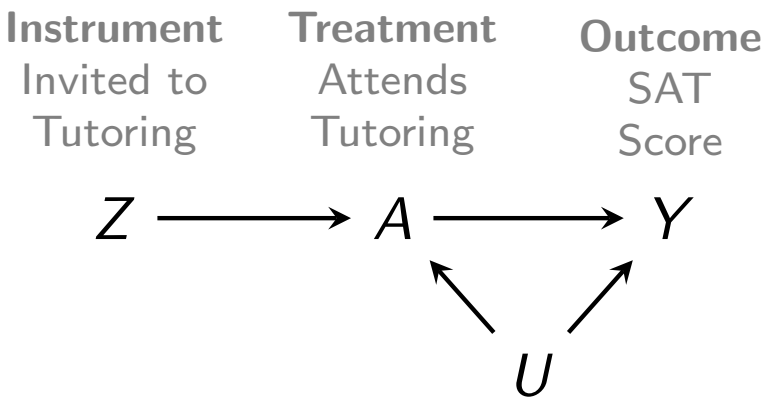
# Requirement 3) Relevance




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$ — hardly any!
3) Divide (1) / (2) — tiny denominator!

# Requirement 4) Monotonicity

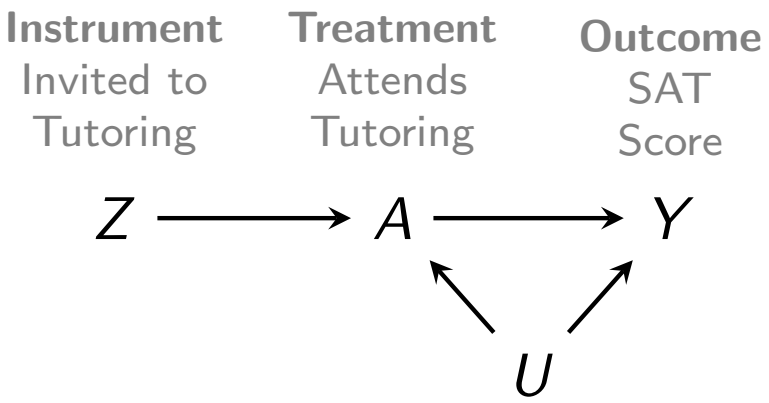



The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Requirement 4) Monotonicity

Assume the invitation never **prevents** tutoring. $A^{z=1} \geq A^{z=0}$



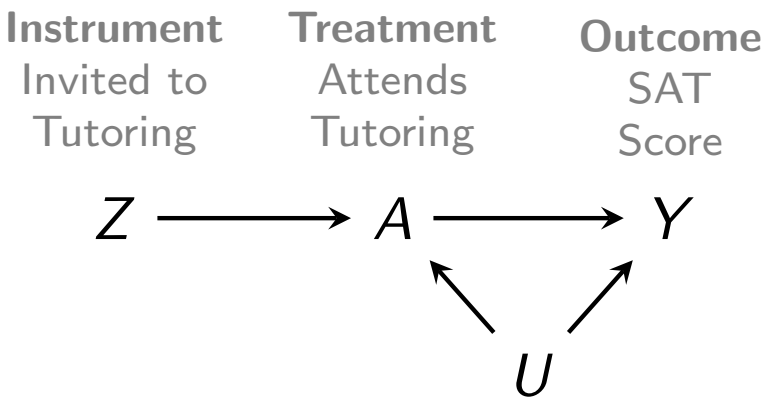

The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Requirement 4) Monotonicity

Assume the invitation never **prevents** tutoring. $A^{z=1} \geq A^{z=0}$

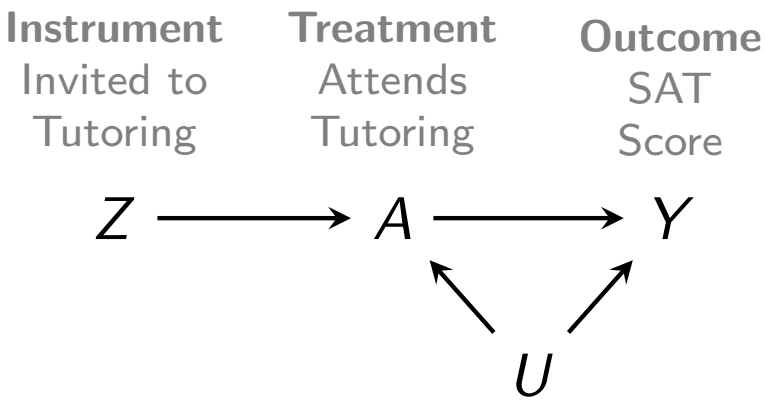
Average effect among the compliers: $\mathrm{E}(Y^{a=1} - Y^{a=0} \mid Z^{a=1} = 1, Z^{a=0} = 0)$




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# What we want




The instrumental variable steps

1) Estimate effect of $Z$ on $Y$
2) Estimate effect of $Z$ on $A$
3) Divide (1) / (2)

# Two stage least squares

When instrument is continuous (or if you have multiple instruments), the Wald estimator doesn't work

- Use $Z_i$ to predict $\hat{A}_i$
  - Exclusion: the variation in $\hat{A}_i$ is only because of $Z_i$ and not $U_i$
- Regress $Y_i$ onto $\hat{A}_i$
- Linear coefficient is the estimate of $\hat{A} \rightarrow Y$

# Two stage least squares

When instrument is continuous (or if you have multiple instruments), the Wald estimator doesn't work

- Use $Z_i$ to predict $\hat{A}_i$
  - Exclusion: the variation in $\hat{A}_i$ is only because of $Z_i$ and not $U_i$
- Regress $Y_i$ onto $\hat{A}_i$
- Linear coefficient is the estimate of $\hat{A} \rightarrow Y$

Same as Wald estimate if instrument and treatment are binary

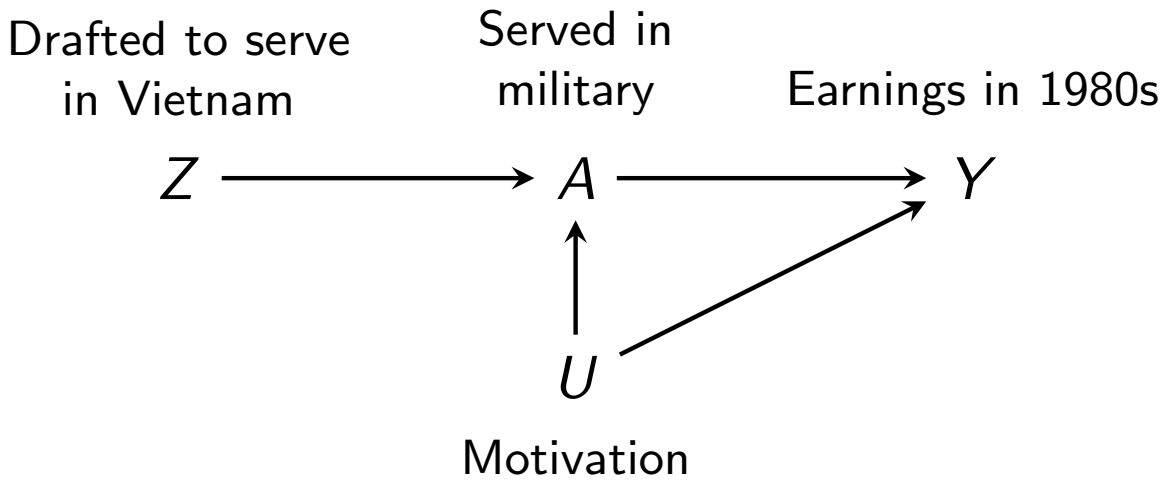
An observational settings that is **clean**

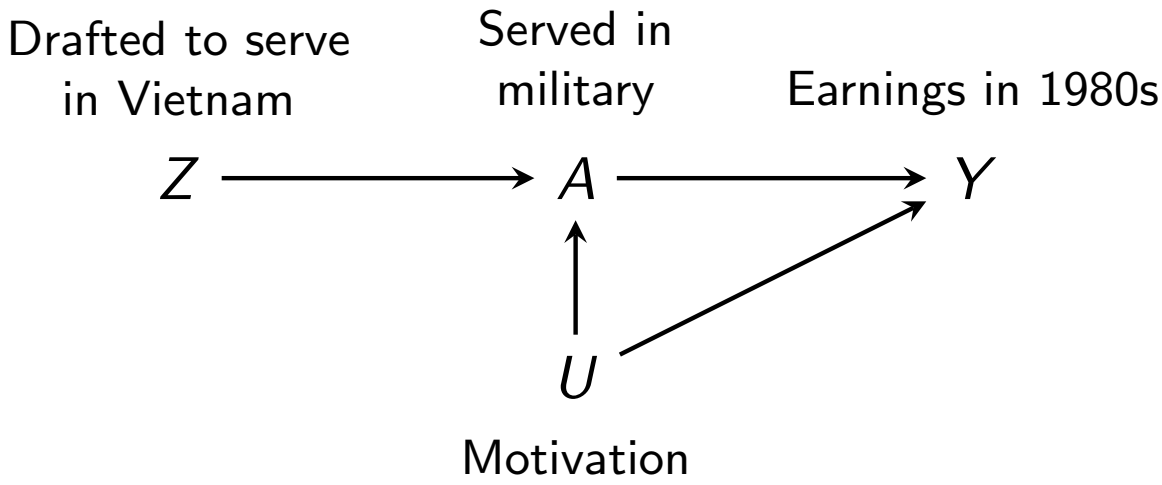
1970 RANDOM SELECTION SEQUENCE, BY MONTH AND DAY

| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 305 | 086 | 108 | 032 | 330 | 249 | 093 | 111 | 225 | 359 | 019 | 129 |
| 2 | 159 | 144 | 029 | 271 | 298 | 228 | 350 | 045 | 161 | 125 | 034 | 328 |
| 3 | 251 | 297 | 267 | 083 | 040 | 301 | 115 | 261 | 049 | 244 | 348 | 157 |
| 4 | 215 | 210 | 275 | 081 | 276 | 020 | 279 | 145 | 232 | 202 | 266 | 165 |
| 5 | 101 | 214 | 293 | 269 | 364 | 028 | 188 | 054 | 082 | 024 | 310 | 056 |
| 6 | 224 | 347 | 139 | 253 | 155 | 110 | 327 | 114 | 006 | 087 | 076 | 010 |
| 7 | 306 | 091 | 122 | 147 | 035 | 085 | 050 | 168 | 008 | 234 | 051 | 012 |
| 8 | 199 | 181 | 213 | 312 | 321 | 366 | 013 | 048 | 184 | 283 | 097 | 105 |
| 9 | 194 | 338 | 317 | 219 | 197 | 335 | 277 | 106 | 263 | 342 | 080 | 043 |
| 10 | 325 | 216 | 323 | 218 | 065 | 206 | 284 | 021 | 071 | 220 | 282 | 041 |
| 11 | 329 | 150 | 136 | 014 | 037 | 134 | 248 | 324 | 158 | 237 | 046 | 039 |
| 12 | 221 | 068 | 300 | 346 | 133 | 272 | 015 | 142 | 242 | 072 | 066 | 314 |
| 13 | 318 | 152 | 259 | 124 | 295 | 069 | 042 | 307 | 175 | 138 | 126 | 163 |
| 14 | 238 | 004 | 354 | 231 | 178 | 356 | 331 | 198 | 001 | 294 | 127 | 026 |
| 15 | 017 | 089 | 169 | 273 | 130 | 180 | 322 | 102 | 113 | 171 | 131 | 320 |
| 16 | 121 | 212 | 166 | 148 | 055 | 274 | 120 | 044 | 207 | 254 | 107 | 096 |
| 17 | 235 | 189 | 033 | 260 | 112 | 073 | 098 | 154 | 255 | 288 | 143 | 304 |
| 18 | 140 | 292 | 332 | 090 | 278 | 341 | 190 | 141 | 246 | 005 | 146 | 128 |
| 19 | 058 | 025 | 200 | 336 | 075 | 104 | 227 | 311 | 177 | 241 | 203 | 240 |
| 20 | 280 | 302 | 239 | 345 | 183 | 360 | 187 | 344 | 063 | 192 | 185 | 135 |
| 21 | 186 | 363 | 334 | 062 | 250 | 060 | 027 | 291 | 204 | 243 | 156 | 070 |
| 22 | 337 | 290 | 265 | 316 | 326 | 247 | 153 | 339 | 160 | 117 | 009 | 053 |
| 23 | 118 | 057 | 256 | 252 | 319 | 109 | 172 | 116 | 119 | 201 | 182 | 162 |
| 24 | 059 | 236 | 258 | 002 | 031 | 358 | 023 | 036 | 195 | 196 | 230 | 095 |
| 25 | 052 | 179 | 343 | 351 | 361 | 137 | 067 | 286 | 149 | 176 | 132 | 084 |
| 26 | 092 | 365 | 170 | 340 | 357 | 022 | 303 | 245 | 018 | 007 | 309 | 173 |
| 27 | 355 | 205 | 268 | 074 | 296 | 064 | 289 | 352 | 233 | 264 | 047 | 078 |
| 28 | 077 | 299 | 223 | 262 | 308 | 222 | 088 | 167 | 257 | 094 | 281 | 123 |
| 29 | 349 | 285 | 362 | 191 | 226 | 353 | 270 | 061 | 151 | 229 | 099 | 016 |
| 30 | 164 | --- | 217 | 208 | 103 | 209 | 287 | 333 | 315 | 038 | 174 | 003 |
| 31 | 211 | --- | 030 | --- | 313 | --- | 193 | 011 | --- | 079 | --- | 100 |

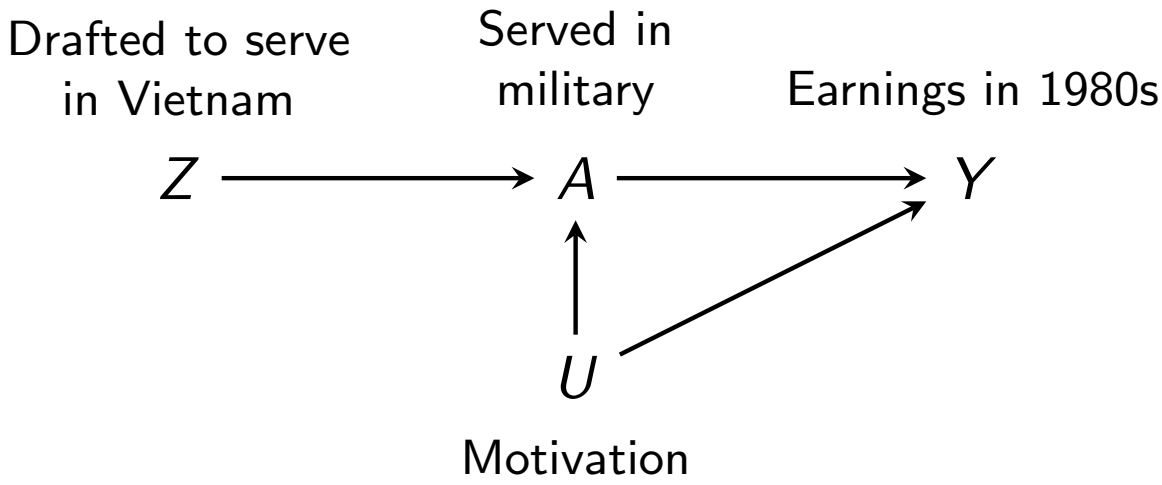
https://www.historynet.com/whats-your-number/

https://commons.wikimedia.org/wiki/File:1969_draft_lottery_photo.jpg


Drafted to serve in Vietnam
Served in military
Earnings in 1980s
Z
A
Y
U
Motivation

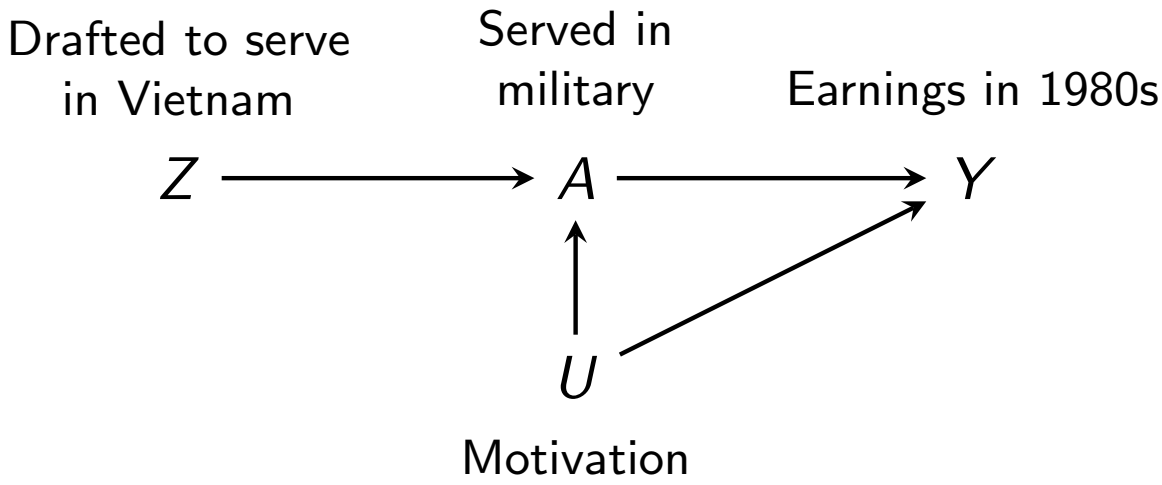



Resulting estimate: Army service causes approximately 2000 loss of income.



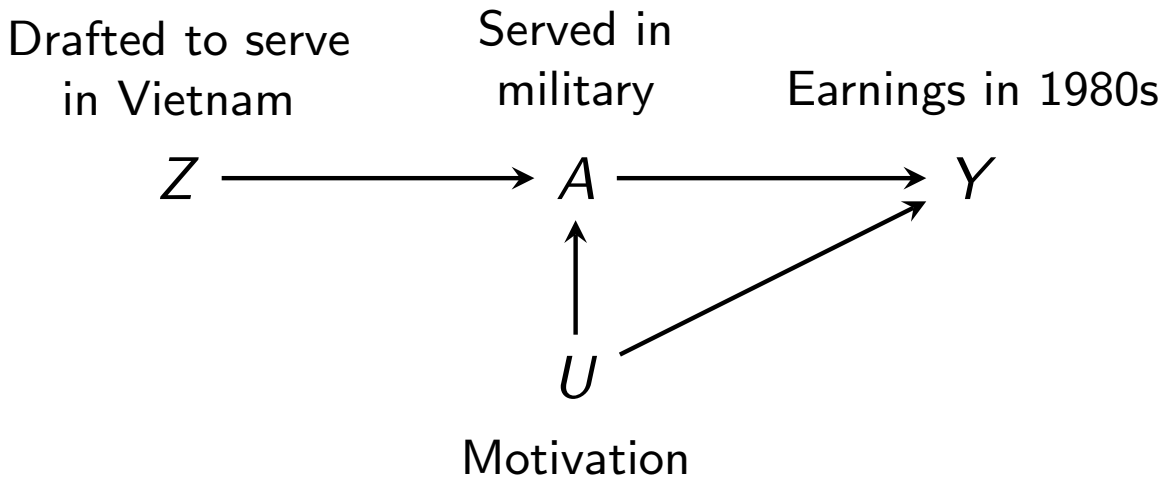

Resulting estimate: Army service causes approximately 2000 loss of income.

- Exogeneity: Randomly selected draft numbers




Resulting estimate: Army service causes approximately 2000 loss of income.

- Exogeneity: Randomly selected draft numbers
- Exclusion: Being drafted affects earnings only through service




Resulting estimate: Army service causes approximately 2000 loss of income.

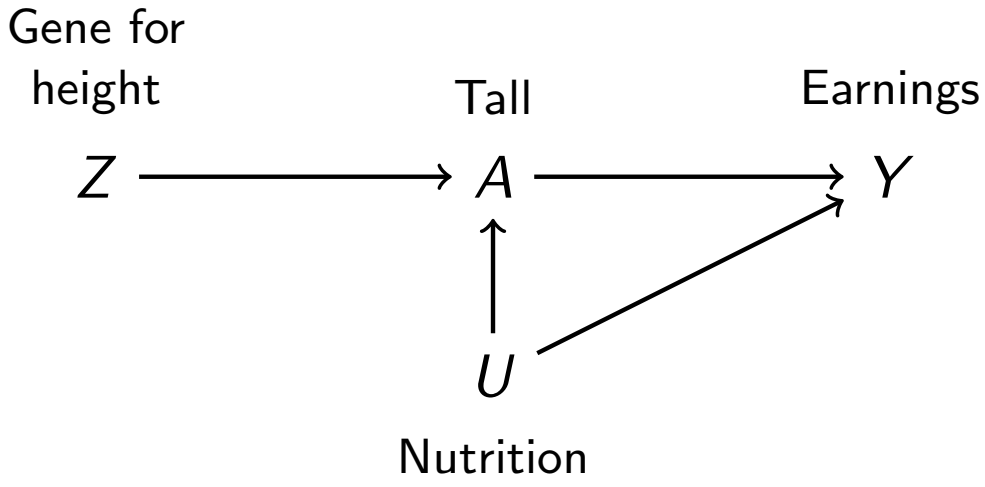
- Exogeneity: Randomly selected draft numbers
- Exclusion: Being drafted affects earnings only through service
- Relevance: Being drafted has a causal effect on serving in the military
- Monotonicity: No one joins the military in defiance of not being drafted

Angrist, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records. The American Economic Review, 313-336.

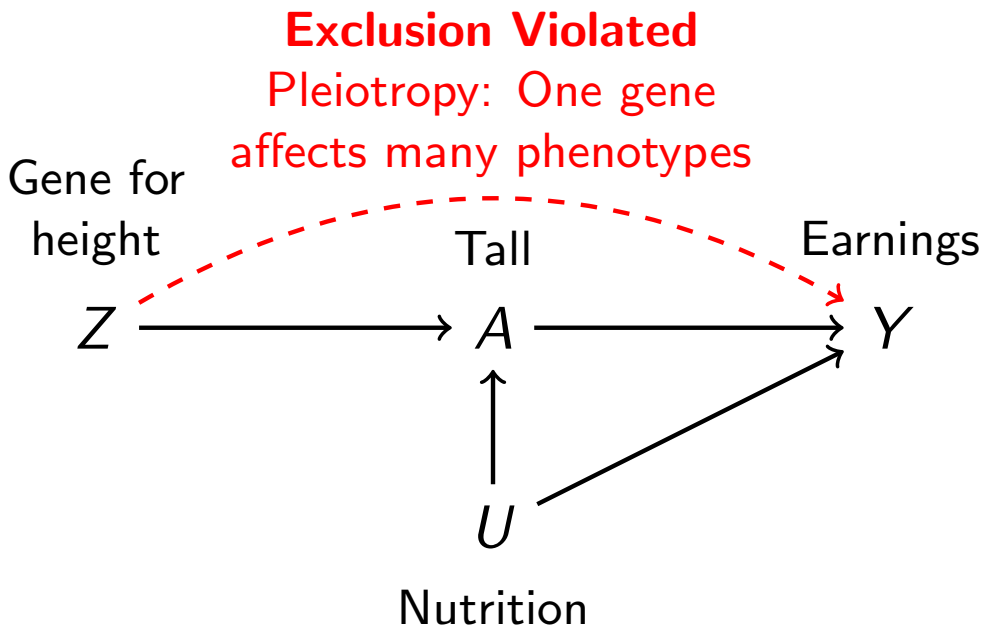
Observational settings that are **less straightforward**

# A difficult example: Genes as instruments

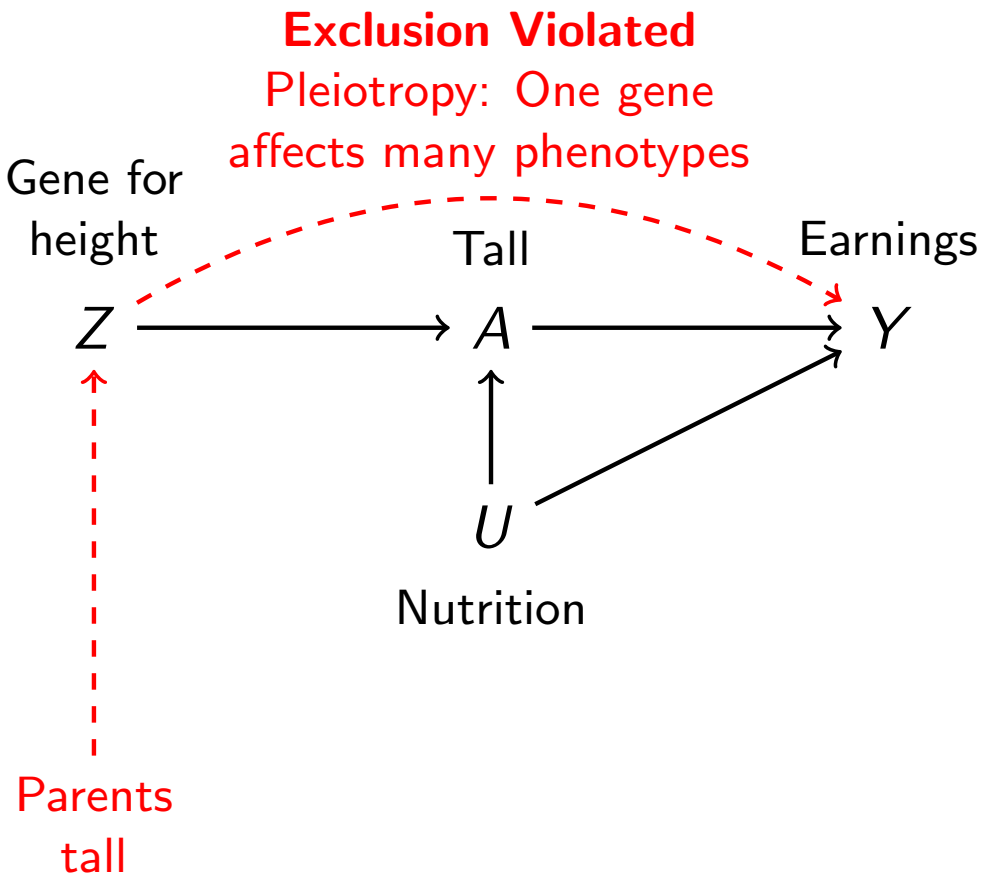
# A difficult example: Genes as instruments



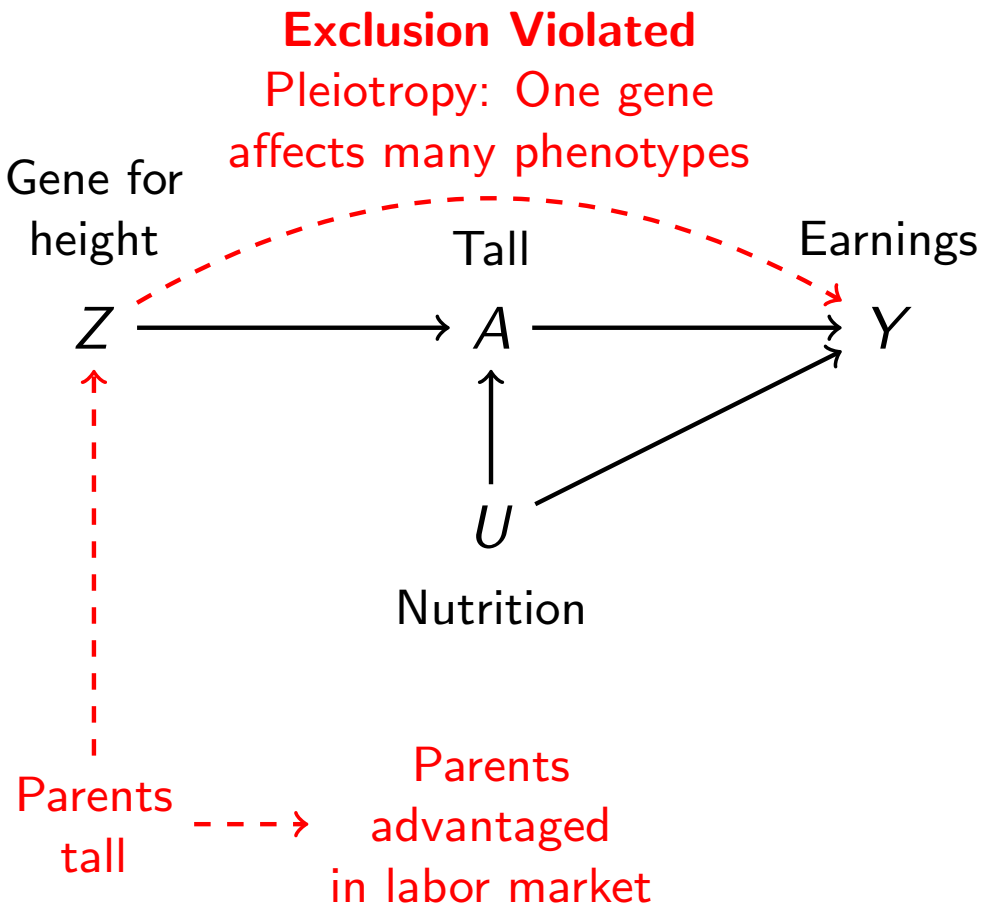
# A difficult example: Genes as instruments



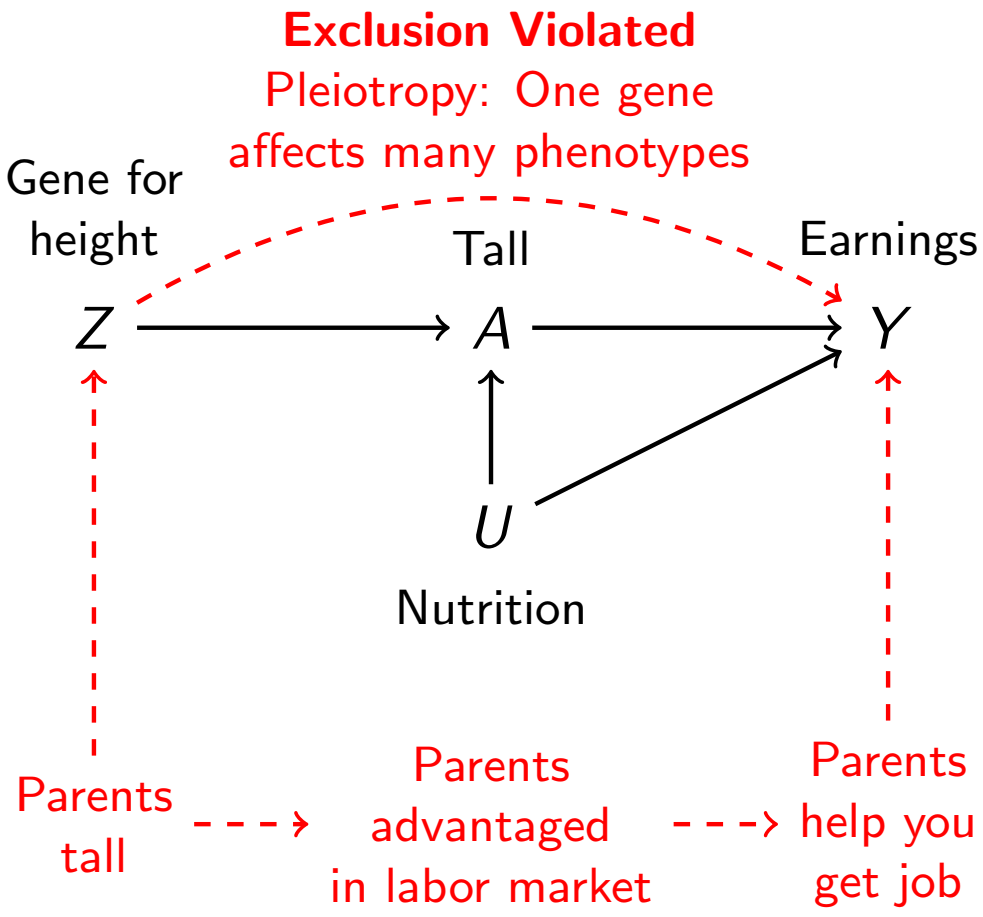
# A difficult example: Genes as instruments



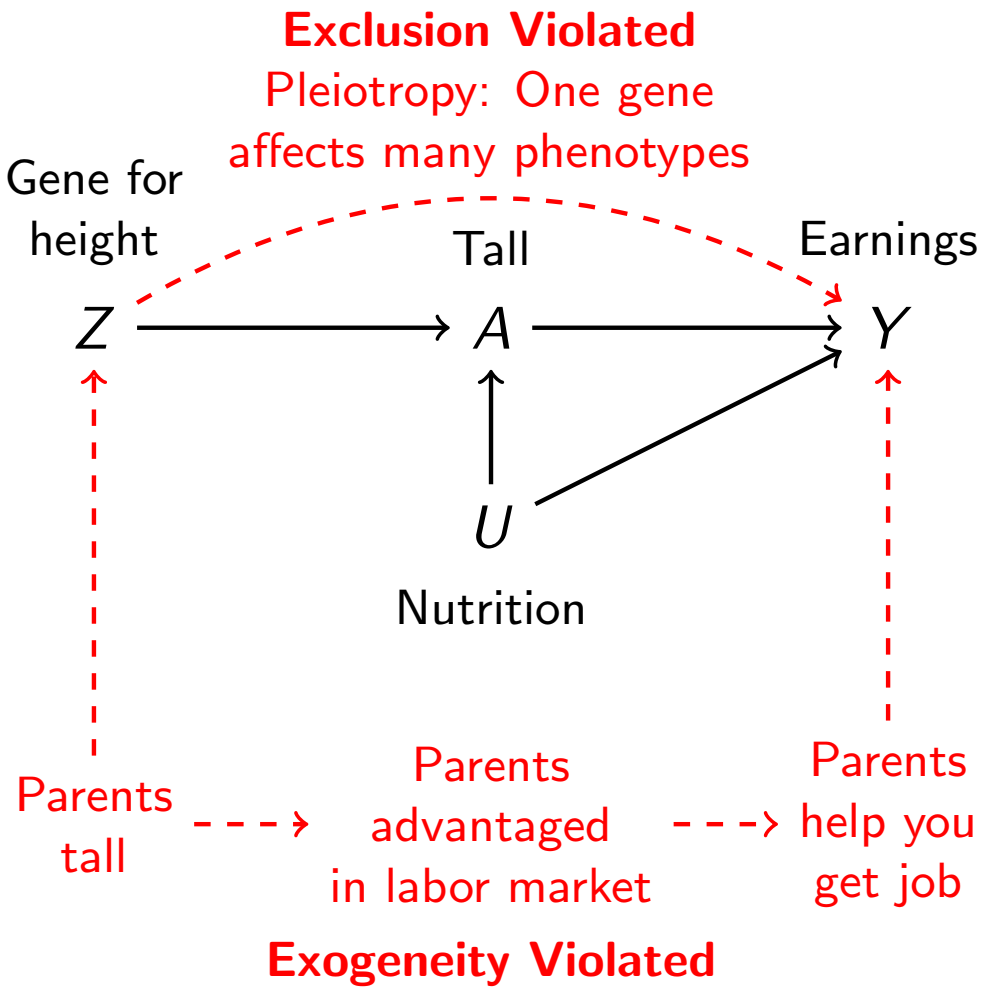
# A difficult example: Genes as instruments



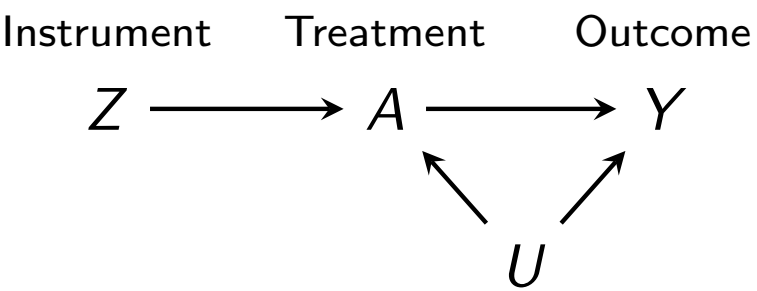
# A difficult example: Genes as instruments



# A difficult example: Genes as instruments



# Learning goals for today

At the end of class, you will be able to:

1. Explain requirements for IV in observational settings
   - instrument exchangeability: $Z \to Y$ is identified
   - no direct effect: $Z \to Y$ is entirely $Z \to A \to Y$
   - relevance: $Z \to A$ is sufficiently large
   - monotonicty: instrument affects treatment in one direction
2. Argue for when these hold or do not hold