



SUPPRESSION,  
COLLIDER UND  
CONFOUNDER

# ÜBERSICHT

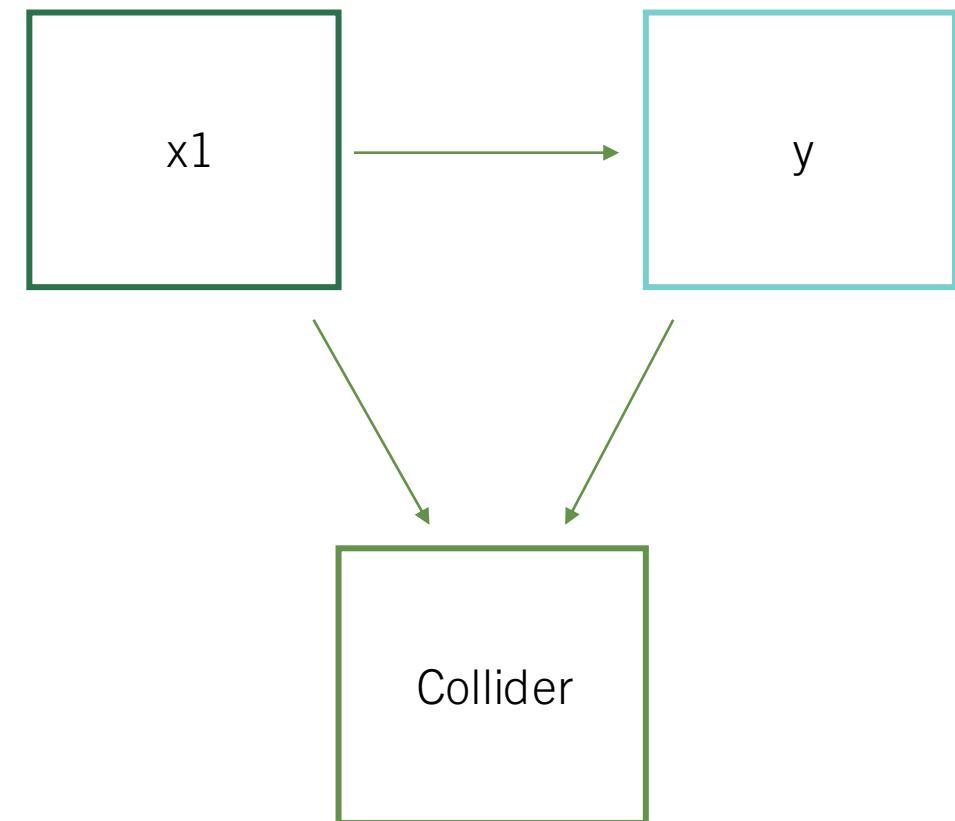
---

- Definition von Collider, Confounder und Suppression
- Unterschiede zwischen Suppression, Collidern und Confoundern
- Wie erkennt man sie?
  - Anwendung von DAGs
- Warum ist es wichtig, sie zu erkennen?
- Benutzte Literatur

# WAS IST EIN COLLIDER?

---

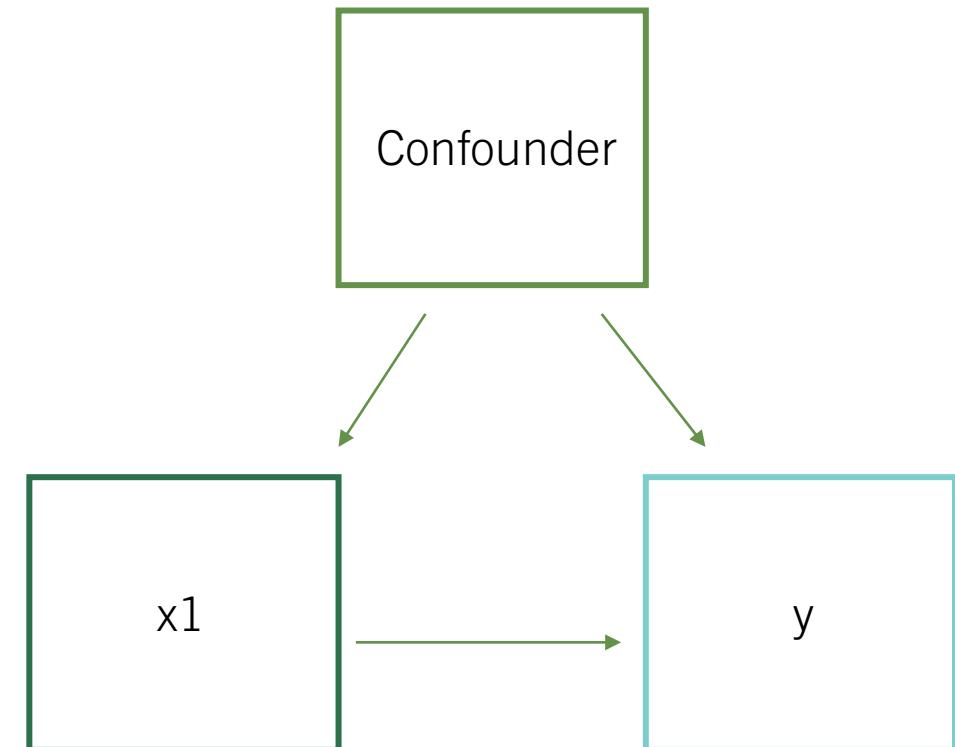
- Collider wird unabhängig voneinander von x (UV) und y (AV) beeinflusst
- Man darf nicht für Collider kontrollieren, da kausaler Effekt sonst verzerrt geschätzt wird
- Expliziter Collider: Collider wird gemessen und für ihn kontrolliert
- Impliziter Collider: wenn Collider zwar nicht gemessen wird, aber die Stichprobe basierend auf seinen Merkmalen ausgewählt wird (z.B. Selection bias, missing data)



# WAS IST EIN CONFOUNDER?

---

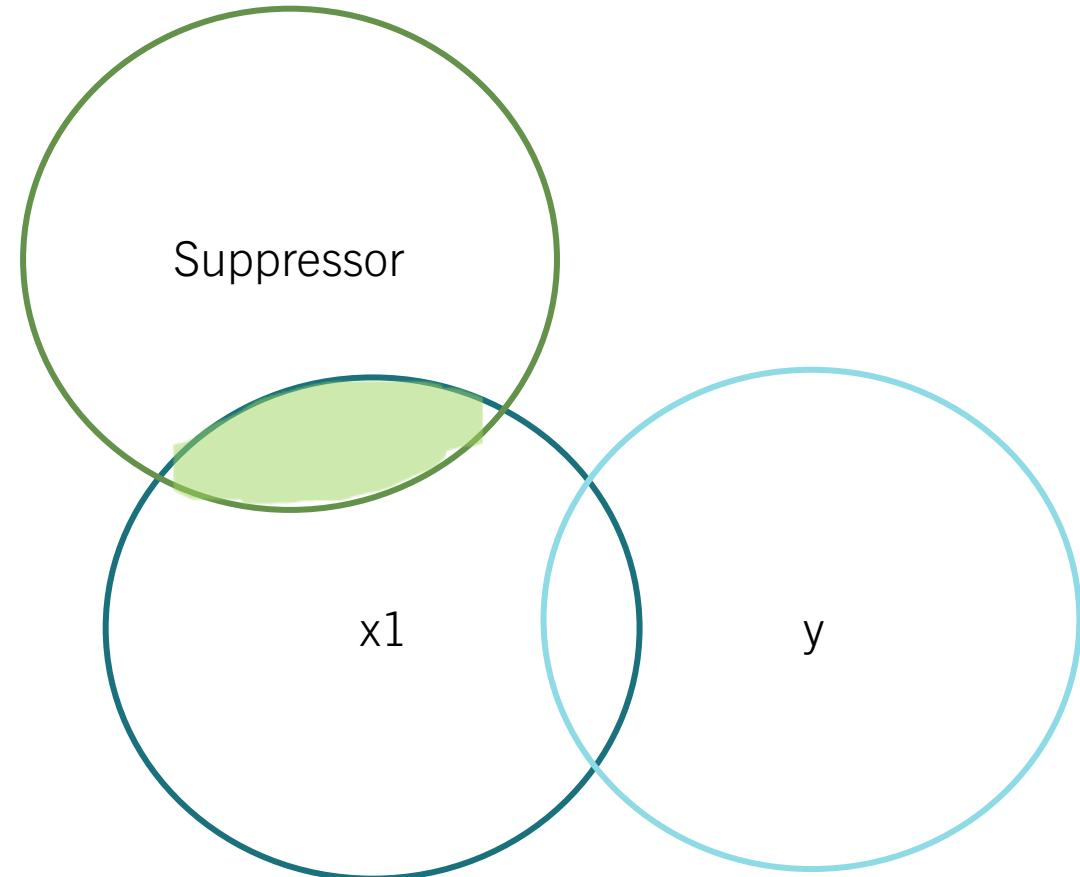
- Drittvariable, die sowohl auf UV als auch auf AV kausalen Einfluss hat
- Ergebnisse zeigen nicht den echten Zusammenhang zwischen AV und UV, weil von Confounder-Variable beeinflusst
- Beeinflusst, ob man bei Zusammenhang von zwei Variablen von Kausalität sprechen kann -> nur möglich, wenn alle relevanten Confounder ausgeschlossen wurden
- Confounder haben kausalen Einfluss auf UV und AV, daher muss für ihn kontrolliert werden, um kausalen Effekt von UV auf AV zu erkennen



# WAS IST SUPPRESSION?

---

- Variable, die Varianz von einer UV, die nicht mit der AV korreliert, bindet und dadurch die Vorhersagekraft dieser UV erhöht
- Dadurch Verstärkung des Zusammenhangs zwischen UV und AV
- Kontraintuitiv, weil man eigentlich annimmt, dass durch die Kontrolle einer zusätzlichen Variable in einem Modell der Zusammenhang zwischen der UV und der AV schwächer wird
- Kommen oft vor, wenn es hohe Korrelationen zwischen den Prädiktoren gibt



# TYPEN VON SUPPRESSION

---

- Absolute Suppression: Standardisierter Regressionskoeffizient des Prädiktors, auf den der Suppressor Einfluss nimmt, wird größer als seine bivariate Korrelation mit dem Kriterium
- Klassische Suppression: Die Suppressorvariable korreliert nicht mit dem Kriterium -> restriktive Version einer absoluten Suppression
- Negative Suppression: Prädiktor, auf den der Suppressor Einfluss nimmt, hat einen negativen Regressionskoeffizienten bei positiver Korrelation mit dem Kriterium
- Reziproke Suppression: Prädiktoren sind gleichzeitig Suppressor und Variable, auf die Suppressor wirkt -> beide Regressionskoeffizienten sind höher als ihre jeweilige Korrelation mit dem Kriterium, oder haben entgegengesetztes Vorzeichen

# UNTERSCHIEDE ZWISCHEN SUPPRESSION, COLLIDERN UND CONFOUNDERN

---

- Alle sind Drittvariablen, die einen Einfluss auf den Zusammenhang zwischen UV und AV haben
- Confounder und Collider sind gegensätzlich
- Für Confounder sollte kontrolliert werden, für Collider nicht; bei Suppressoren ist es eine inhaltliche Entscheidung
- Suppression verstärkt den Zusammenhang zwischen UV und AV durch Kontrolle, anstatt von schwächerem Zusammenhang (Confounder)

# WIE ERKENNT MAN COLLIDER UND CONFOUNDER?

---

- Directed Acyclic Graphs (DAG) können verwendet werden, um Confounder und Collider zu identifizieren
- DAG: formale Darstellung kausaler Zusammenhänge in graphischer Form
- Verbildlicht die Annahmen über die Zusammenhänge zwischen Variablen
  - Variablen werden als „Knoten“ dargestellt
  - Verbindungen werden als Linien dargestellt
- Acyclic bedeutet, dass es keine feedback-loops gibt, eine Variable kann also nicht ihr eigener Nachkomme sein

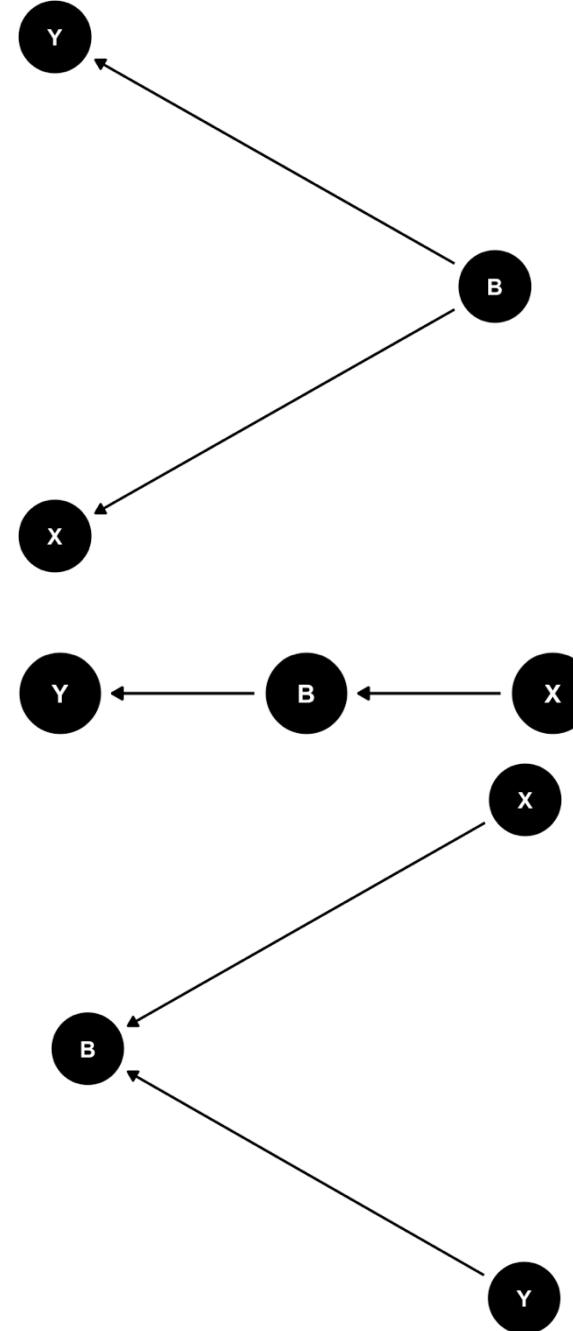
# DAG: MATHEMATISCHE BASIS

---

- Dienen als nicht-parametrische Strukturgleichungsmodelle
- Form des Zusammenhangs zwischen Variablen ist irrelevant
  - Egal, welcher Zusammenhang konkret vorliegt (quadratisch, linear, ...)
  - Nur die Richtung der Pfeile ist von Bedeutung

# ARTEN VON PFADEN

- Fork
  - Knoten, der zwei Variablen verursacht, wobei B ein Confounder ist
- Chains
  - Direkter kausaler Pfad, wobei B einen Mediator darstellt
- Inverted Forks
  - Knoten, auf den zwei oder mehr Pfeile zeigen, wobei B einen Collider darstellt



# BACK-DOOR KRITERIUM

---

- $Z$  = Set von Variablen, die zur Kontrolle mit herangezogen wird
- $X$  = Unabhängige Variable
- $Y$  = Abhängige Variable
- Das Set von Variablen  $Z$  erfüllt die Kriterien relativ zu  $X$  &  $Y$ , wenn es zwei Bedingungen einhält
  - Kein Knoten in  $Z$  ein Nachkomme von  $X$  ist
  - $Z$  blockiert jeden Pfad zwischen  $X$  &  $Y$ , der mit einem Pfeil in  $X$  endet

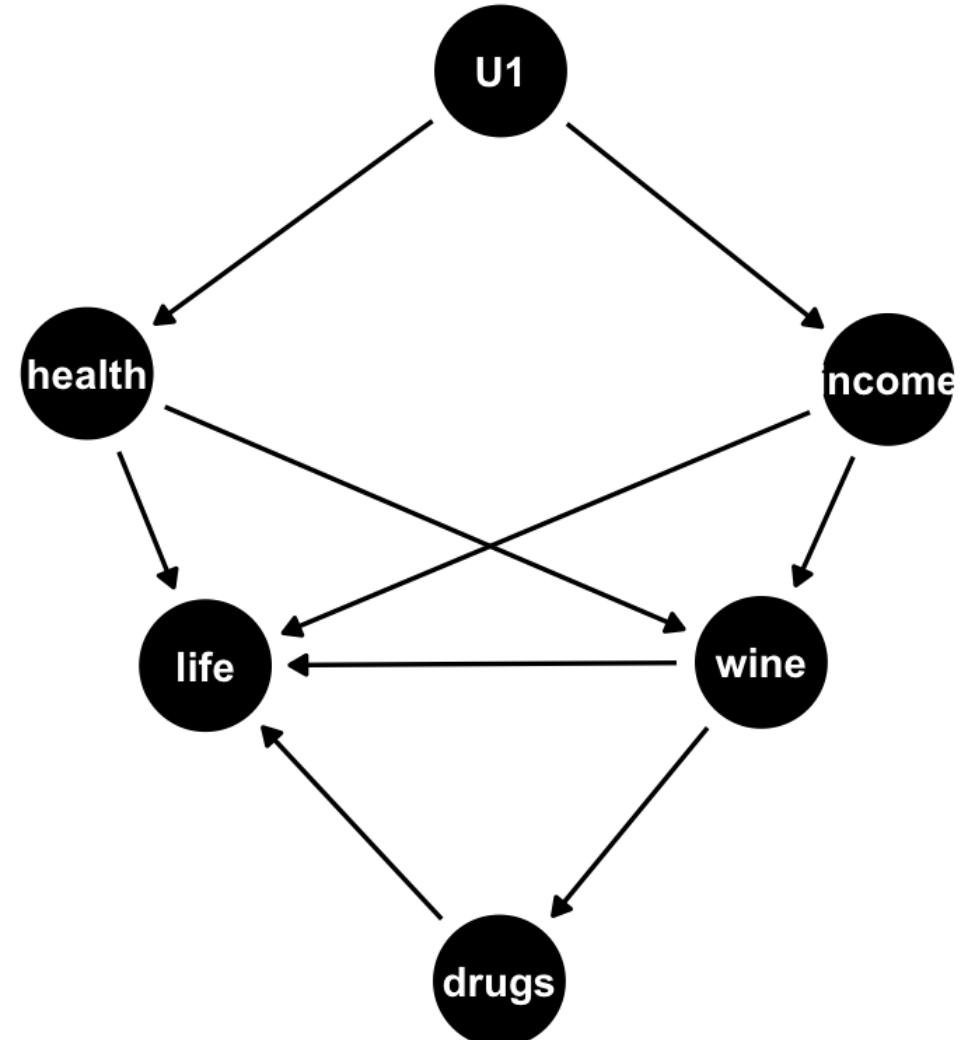
# BACK-DOOR KRITERIUM

---

1. Kausales Modell festlegen
2. Pfade zwischen X und Y identifizieren
3. Back-Door-Pfade erkennen
4. Blockierende Variablen bestimmen
5. Ausschlussregel & Back-Door-Kriterium beachten
6. Kausalen Effekt schätzen

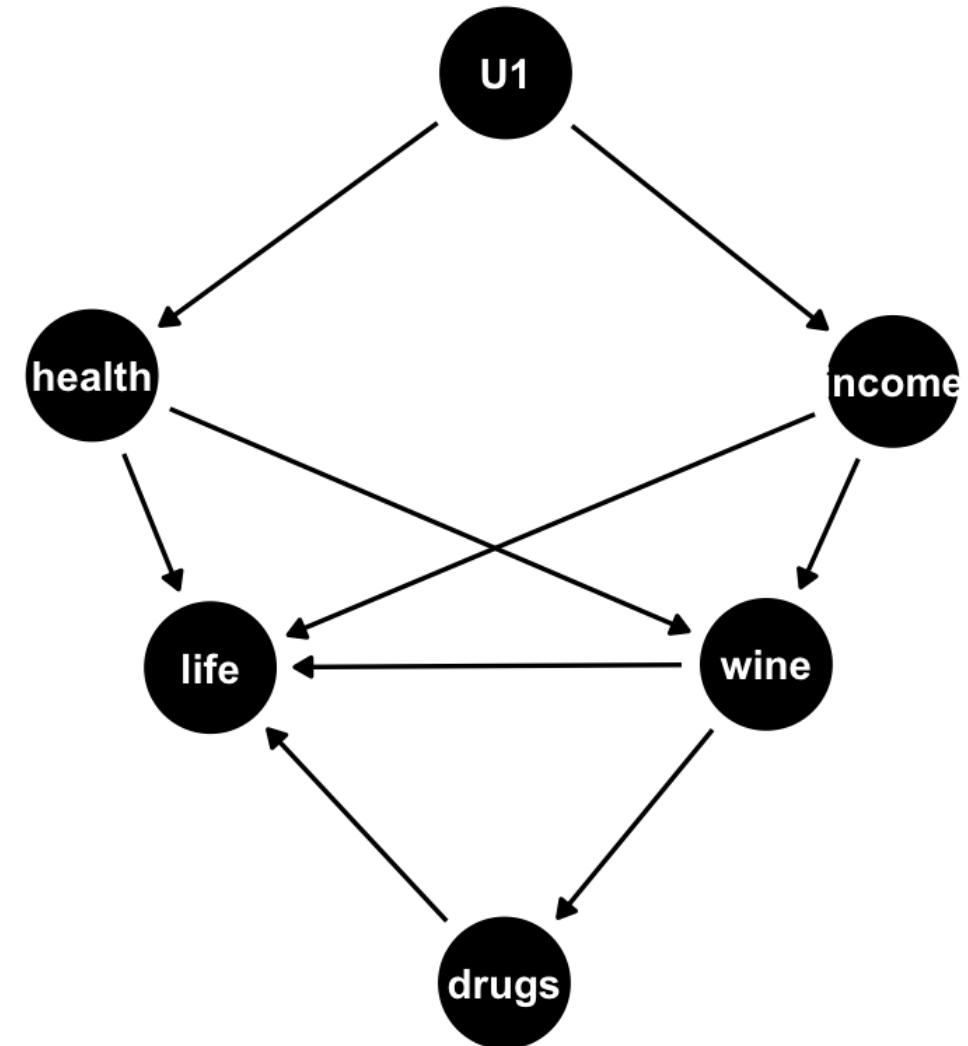
# BEISPIEL R-CODE FÜR EIN DAG

```
# install.packages("ggdag")
# install.packages("ggplot2")
# library(ggdag)
# library(ggplot2)
# theme_set(theme_dag())
# Exp <- dagify(life ~ income + health + wine + drugs,
#               income ~ U1,
#               health ~ U1,
#               wine ~ health + income,
#               drugs ~ wine)
# ggdag(Exp) + coord_flip() + theme_void()
```



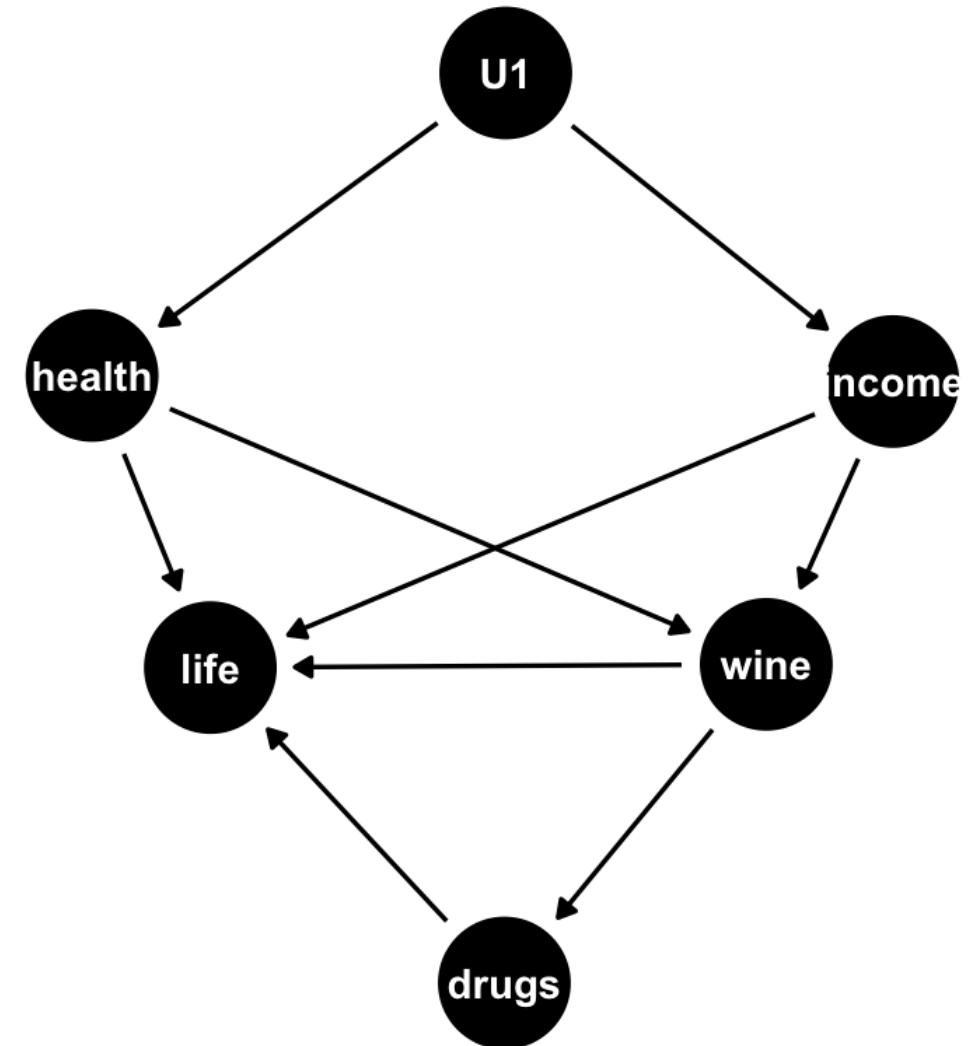
# BEISPIEL: PFADE IM DAG

- Wir interessieren uns für den Einfluss von “wine” auf “life”
- wine  $\rightarrow$  life
- wine  $\rightarrow$  drugs  $\rightarrow$  life
- wine  $<-$  health  $\rightarrow$  life
- wine  $<-$  income  $\rightarrow$  life
- wine  $<-$  health  $<-$  U1  $\rightarrow$  income  $\rightarrow$  life
- wine  $<-$  income  $<-$  U1  $\rightarrow$  health  $\rightarrow$  life



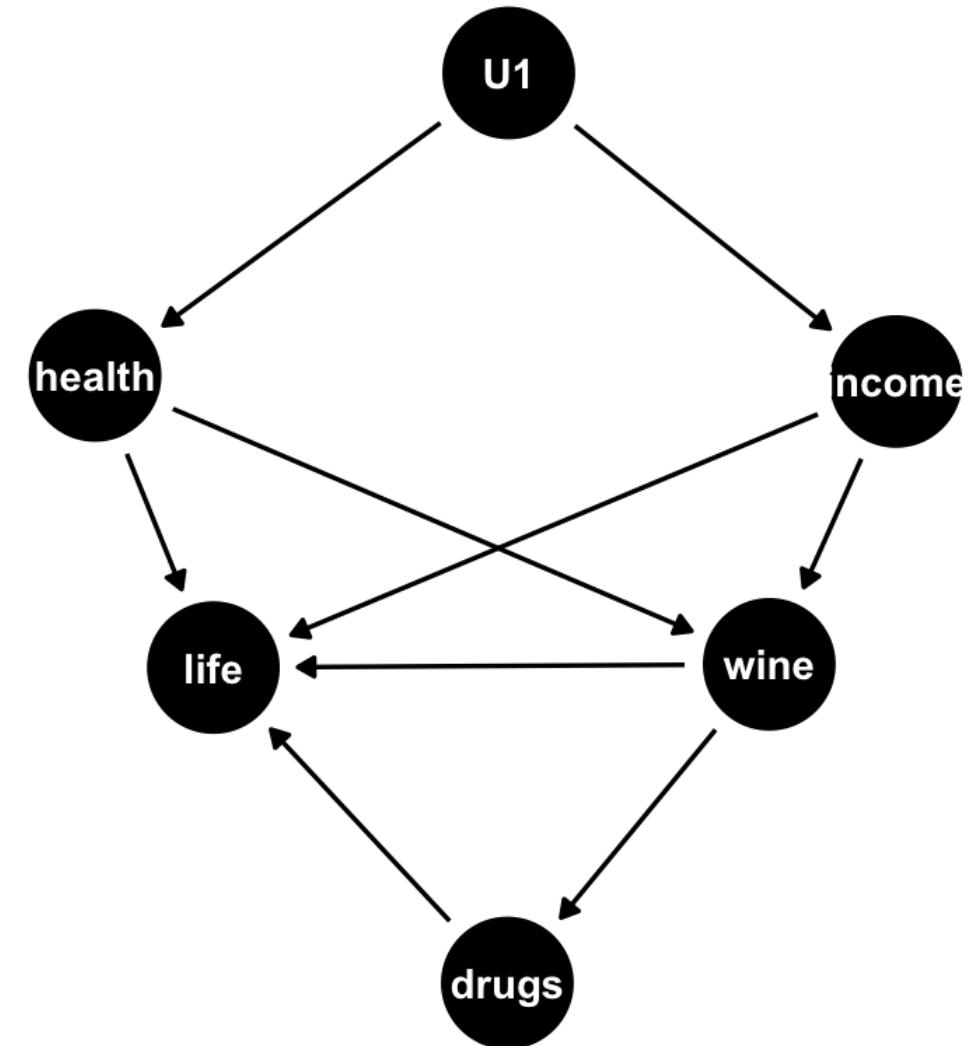
# BEISPIEL: PFADE IM DAG

- Front Door / Back Door
- wine  $\rightarrow$  life
- wine  $\rightarrow$  drugs  $\rightarrow$  life
- wine  $\leftarrow$  health  $\rightarrow$  life
- wine  $\leftarrow$  income  $\rightarrow$  life
- wine  $\leftarrow$  health  $\leftarrow$  U1  $\rightarrow$  income  $\rightarrow$  life
- wine  $\leftarrow$  income  $\leftarrow$  U1  $\rightarrow$  health  $\rightarrow$  life



## BEISPIEL: FÜR “HEALTH” KONTROLIEREN

- Front Door / Open Back Door / Closed Back Door
- wine  $\rightarrow$  life
- wine  $\rightarrow$  drugs  $\rightarrow$  life
- wine  $\leftarrow$  health  $\rightarrow$  life
- wine  $\leftarrow$  income  $\rightarrow$  life
- wine  $\leftarrow$  health  $\leftarrow$  U1  $\rightarrow$  income  $\rightarrow$  life
- wine  $\leftarrow$  income  $\leftarrow$  U1  $\rightarrow$  health  $\rightarrow$  life

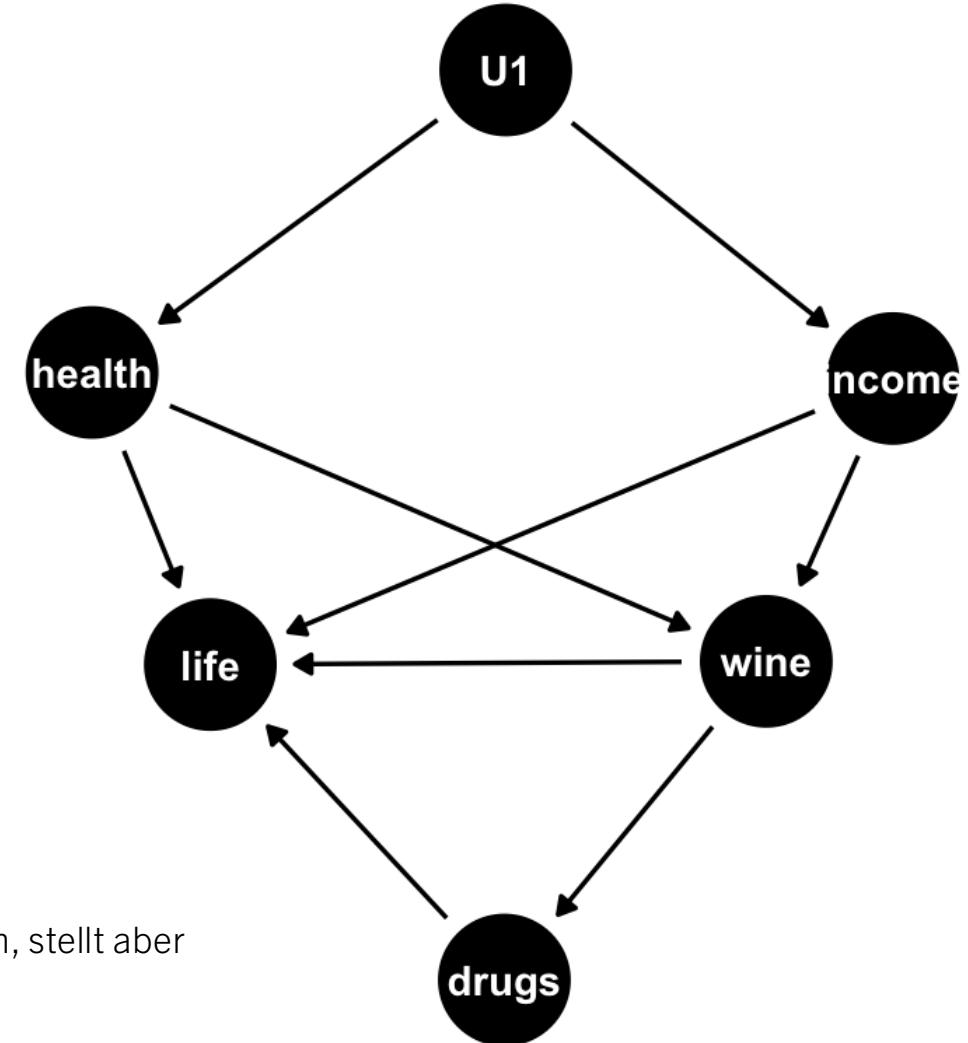


# BEISPIEL: FÜR “HEALTH” UND “INCOME” KONTROLIEREN

- Front Door / Open Back Door / Closed Back Door
- wine  $\rightarrow$  life
- wine  $\rightarrow$  drugs  $\rightarrow$  life
- wine  $<- \text{health} \rightarrow$  life
- wine  $<- \text{income} \rightarrow$  life
- wine  $<- \text{health} <- \text{U1} \rightarrow \text{income} \rightarrow$  life
- wine  $<- \text{income} <- \text{U1} \rightarrow \text{health} \rightarrow$  life

}

Doppelt geschlossen, stellt aber  
kein Problem dar



# WIE ERKENNT MAN SUPPRESSION?

---

- Theoretische Struktur festlegen
- Statistischer Vergleich
  - Zero-order Effekt Modell erstellen
  - Vergleich mit Modell mit extra Variable
  - Wenn der Effekt der UV steigt, liegt Suppression vor
- Ergebnis-Check
  - Signifikanz, Richtung und theoretische Plausibilität prüfen
- Backdoor-Kriterien prüfen

# **WARUM IST ES WICHTIG, SUPPRESSION ZU ERKENNEN?**

---

- Suppressoren sind wichtig, um zu erklären, welche Wirkung die Variablen in einem Modell haben
- Suppressoren sollten auch mit in Modell aufgenommen werden, weil nicht aufgenommene Suppressoren die Vorhersagekraft der Prädiktoren einschränken

# WARUM IST ES WICHTIG, CONFOUNDER ZU ERKENNEN?

---

- Verzerrte Schätzung von Zusammenhängen durch Confounder
- Lösung: Randomisierte Zuordnung von Versuchspersonen in Versuchsgruppen -> dadurch verbesserte interne Validität
- Alternative, falls randomisierte Zuteilung nicht möglich ist (z.B. wegen ethischen Problemen): Statistische Kontrolle von Confoundern durch Aufnahme in das Modell als Kovariate
- Auch möglich: Sensitivitätsanalysen, um zu untersuchen, ob potenzielle Confounder einen starken Effekt auf den Koeffizienten hätten

# WARUM IST ES WICHTIG, COLLIDER ZU ERKENNEN?

---

- Probleme mit Generalisierbarkeit durch implizite Collider
- Verzerrte Schätzungen des Populationseffekts
- Mehrere Lösungen:
  - Sensitivitätsanalysen
  - Vorkehrungen, um Collider zu vermeiden (repräsentative Stichproben)
  - Identifikation durch DAGs, ohne explizit statistisch kontrollieren zu müssen

# BENUTZTE LITERATUR

---

Frank, K. A. (2000). Impact of a Confounding Variable on a Regression Coefficient. *Sociological Methods & Research*, 29(2), 147–194. <https://doi.org/10.1177/0049124100029002001>

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, 14(1), 29–46. <https://doi.org/10.1214/ss/1009211805>

Lamp, S. J., & MacKinnon, D. P. (2025). Correcting for collider effects and sample selection bias in psychological research. *Psychological Methods*, 30(5), 1169–1184. <https://doi.org/10.1037/met0000659>

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2010). Equivalence of the Mediation, Confounding and Suppression Effect. *Prevention science*, 11(4), 173–181.

MacKinnon, D. P., & Lamp, S. J. (2021). A Unification of Mediator, Confounder, and Collider Effects. *Prevention Science*, 22(8), 1185–1193. <https://doi.org/10.1007/s11121-021-01268-x>

Martinez Gutierrez, N., & Cribbie, R. (2021). Incidence and interpretation of statistical suppression in psychological research. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, 53(4), 480–488. <https://doi.org/10.1037/cbs0000267>

Pearl, J. (2013). Causality: Models, Reasoning And Inference (2nd ed). *Cambridge University Press*.  
<https://doi.org/10.1017/CBO9780511803161>

<https://cran.r-project.org/web/packages/ggdag/vignettes/intro-to-dags.html>