



The Book of WHY  
干擾與去干擾

--剷除潛在變項--

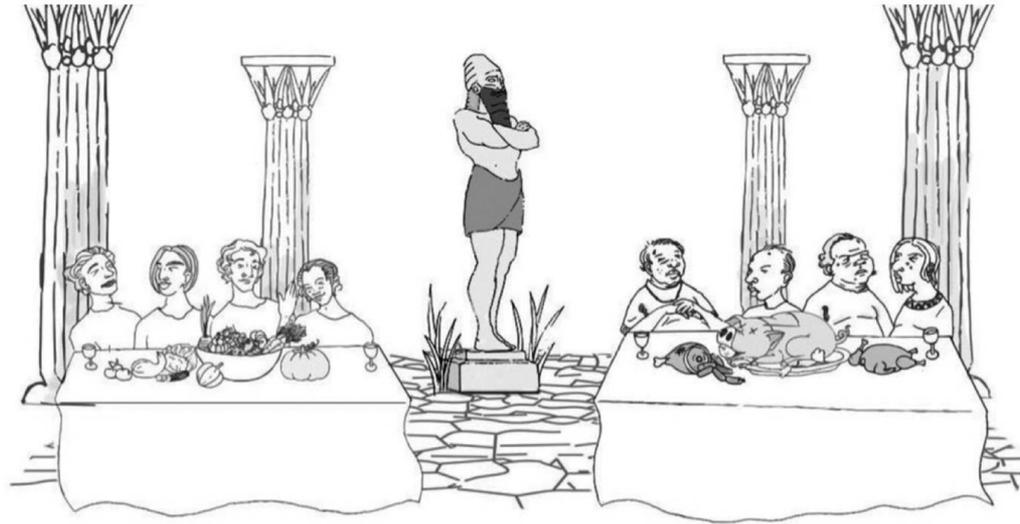
107081025 李念宮

108070015 張為淳

# 概述



# 史上首次對照實驗



**Intervention**  
(Vegetarian Diet)

**Control**  
(King's Diet)

# 對照實驗

## Controlled experiment

### 原理

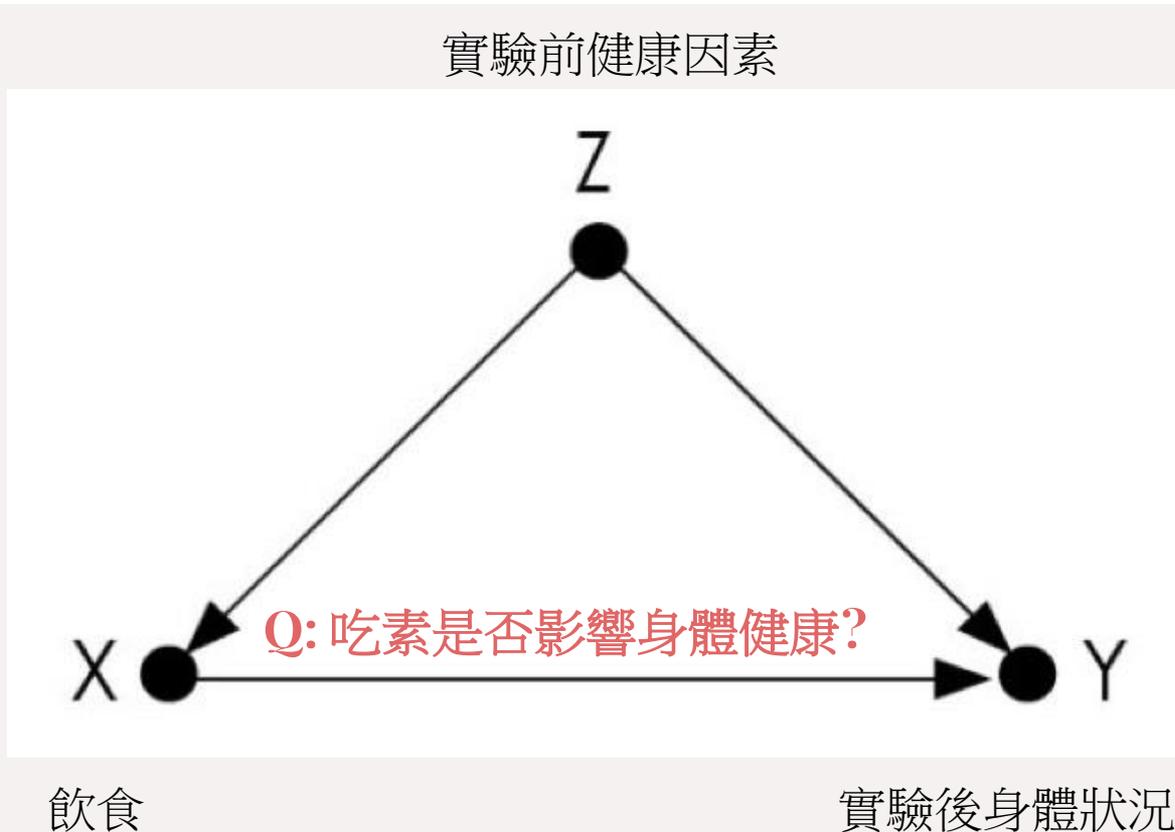
找來兩組條件完全相同的人  
新處理 vs. 舊處理(or 不處理)

觀察差異

然而，健康因素？

# 干擾偏差 confounding bias

X = 原因  
Y = 實驗結果  
Z = 干擾因子



“

干擾 confounding (混合)

→ X與 Y真正的因果效應與

Z在 X與 Y間造成的假相關混合在一起

# 以Z執行調整or對照 Adjusting or controlling for Z

過度輕視  
不談論因果

過度重視  
列出所有干  
擾因子



對干擾因子恐懼

標準去干擾方法：

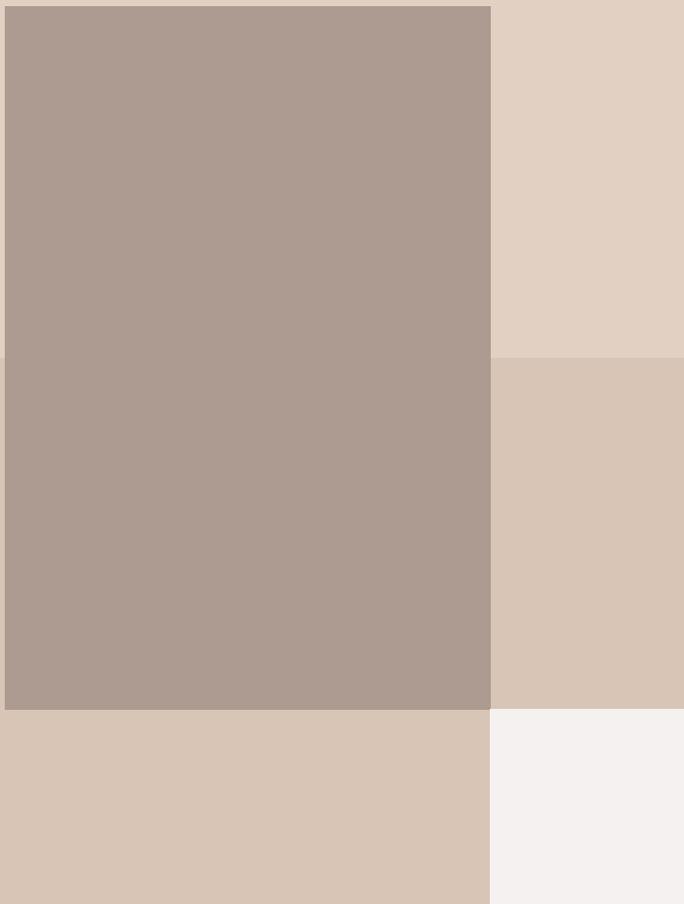
隨機對照試驗

**Random control trial, RCT**

## 目標2.



# 干擾 confounding 的定義



- 敘述性定義：干擾因子是同時與X和Y相關的變項  
\*流行病學定義
- 程序性定義：試圖以統計檢定呈現干擾因子的特性  
→不可崩潰性(noncollapsibility)
- 作者定義：使 $P(Y|X) \neq P(Y|do(X))$ 的因素

# 干擾的流行病學定義

1. Z 在整個群體中與X相關
2. Z在未接受X處理的民眾中與Y相關
3. Z不應該位在X和Y間的因果路徑上

## A History of Epidemiologic Methods and Concepts

Edited by Alfredo Morabia

Birkhäuser Verlag  
Basel · Boston · Berlin

## 干擾的流行病學定義

1. Z 在整個群體中與X相關
2. Z在未接受X處理的民眾中與Y相關
3. Z不應該位在X和Y間的因果路徑上

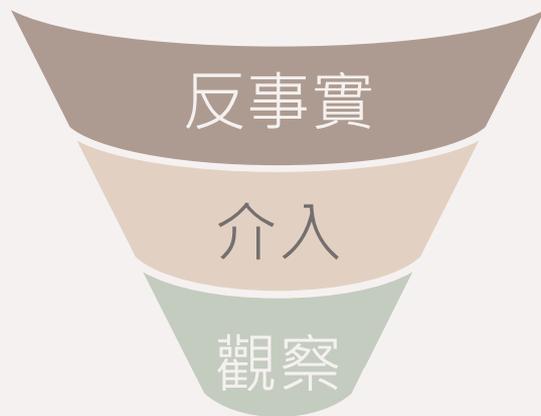
$$(a) X \dashrightarrow Z \dashrightarrow Y$$

$$(b) X \dashrightarrow M \dashrightarrow Y \\ \vdots \\ Z$$

# 可交換性(Exchangeability)

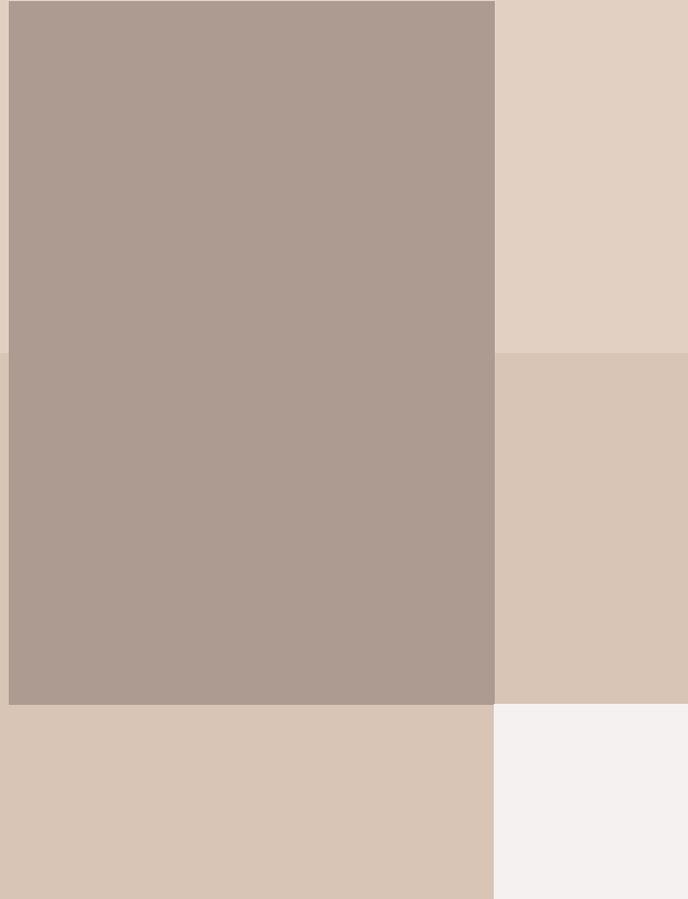
- 回歸對照組( $X = 0$ ) 與處理組( $X = 1$ )
- 以反事實添加變化
- 比較想像與實際的差異
- 沒有差異  $\rightarrow$  沒有干擾

證明敘述性和程序性定義不正確



因果階梯

# 隨機對照試驗 RCT



# Q: 肥料對於產量的影響

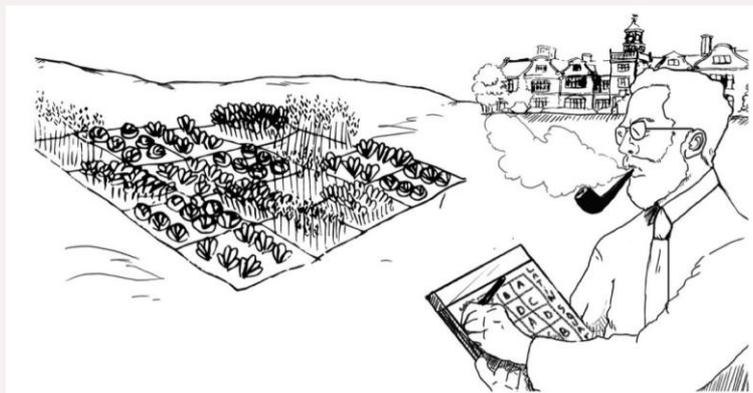
## 費雪拉丁方陣實驗

- 系統化規劃變項
- 確保所有樣本能互相比較

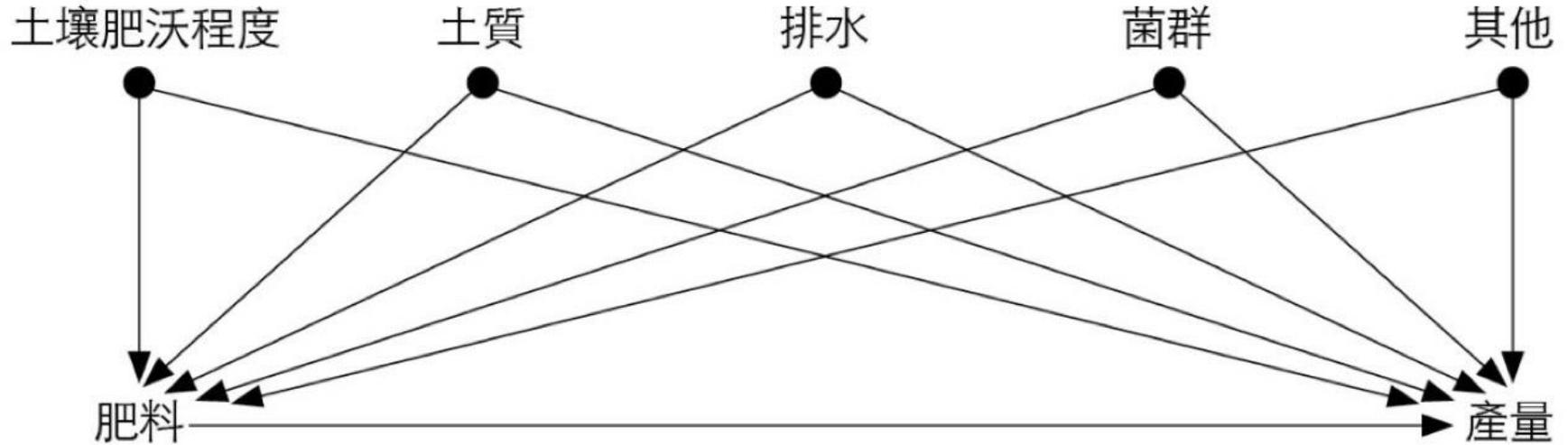
缺點:

1. 無法列出所有干擾因子
2. 沒有效率

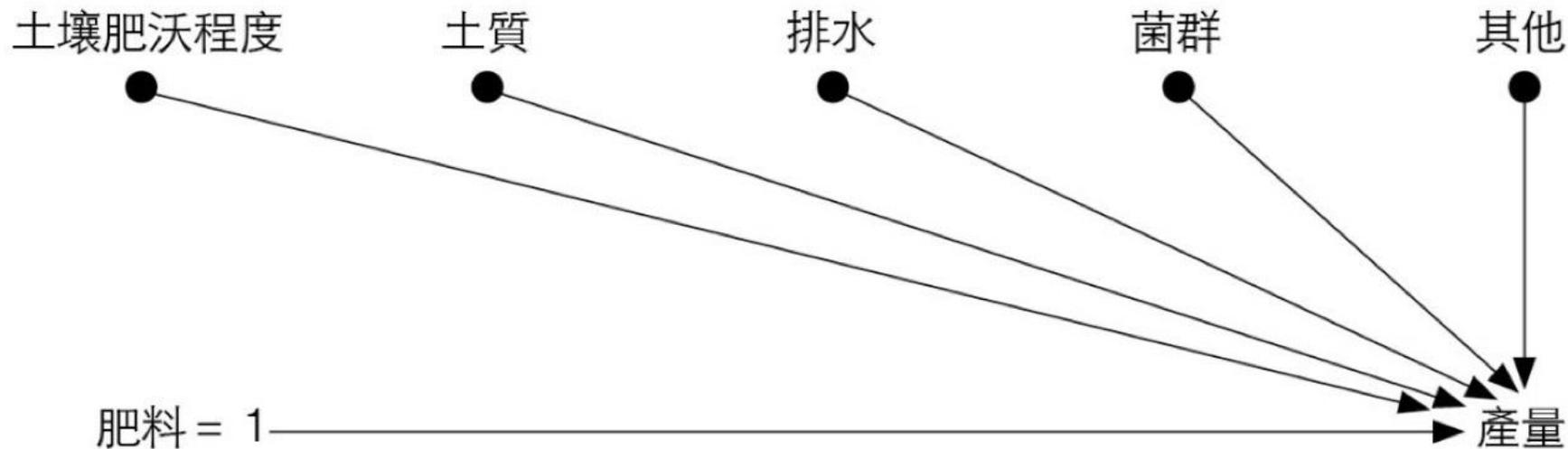
→ 隨機實驗



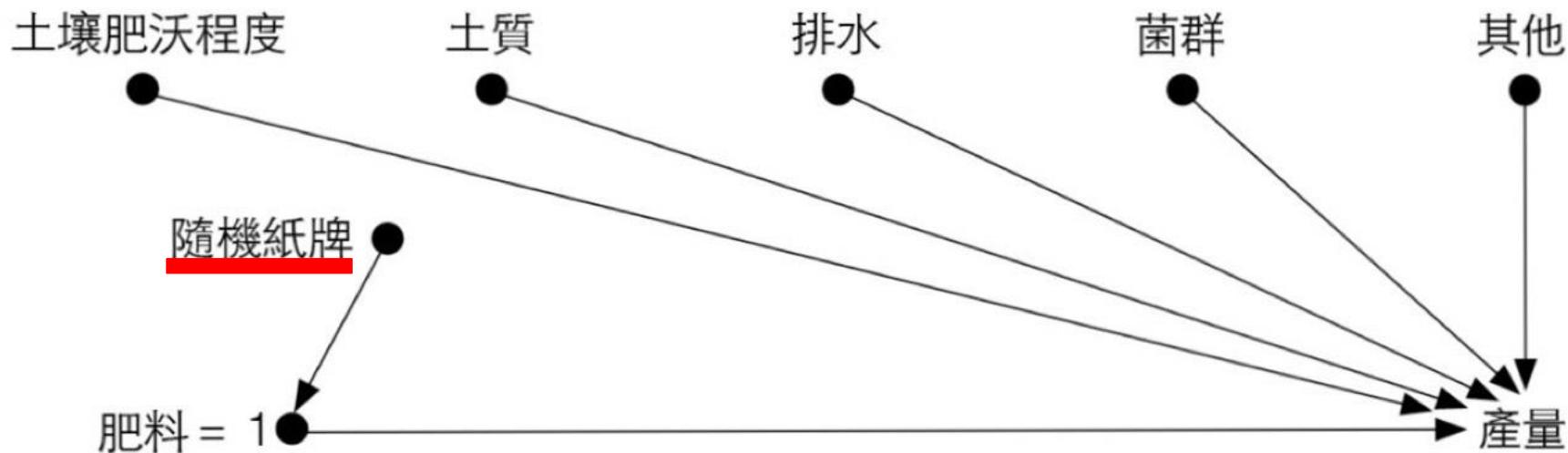
# 模型一—未正確對照的實驗



## 模型二－我們想了解的問題



## 模型三－隨機對照試驗模擬的問題



隨機化模擬了模型二  
→消除原有的干擾因子，不會造成新的干擾因子

# 隨機化

## 優點：

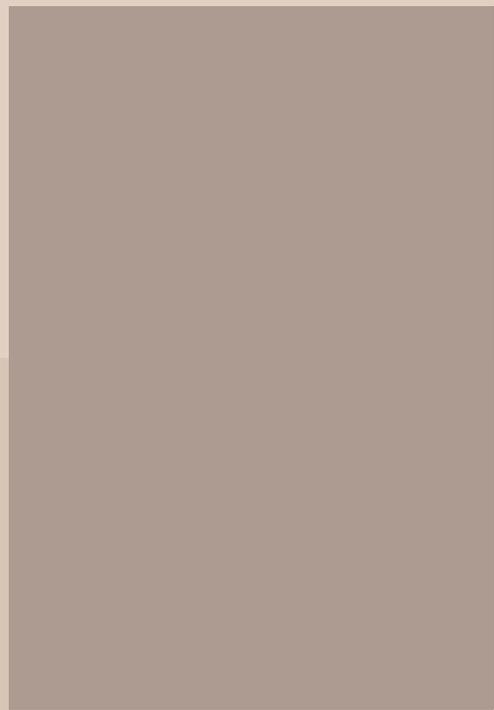
1. 消除干擾偏差
2. 量化不確定性
3. 能切斷所有與隨機化變項間的連結

## 缺點：

1. 某些狀況下無法實際介入
2. 違反研究倫理
3. 難以找到受試者

**RCT主要目標：消除干擾**

# do運算子和後門準則



# do運算子

1. 消除所有指向該變項的箭頭
2. 設定該變項為特定值

作者對干擾的定義：  
使 $P(Y|X) \neq P(Y|do(X))$ 的因素

## 貝氏網路的三種連結(junction)

1.  $A \rightarrow B \rightarrow C$ ，鍊(chain)或中介(mediation)：中介B把關於A的資訊和C隔離(screens off)開來，也把關於C的資訊與A隔離
2.  $A \leftarrow B \rightarrow C$ ，分叉(fork)：B為A和C的共同原因或干擾因子，可使沒有直接因果連結的A和C產生統計相關
3.  $A \rightarrow B \leftarrow C$ ，衝突(collider)：A和C獨立，若將B執行條件化(值設為1)可使A和C變成負相關，稱為衝突偏誤(collider bias)或是自圓其說效應(explain-away effect)

## 以貝氏網路的三種連結執行對照

1.  $A \rightarrow B \rightarrow C$ ：以B執行對照，防止A、C資訊互流
2.  $A \leftarrow B \rightarrow C$ ：以B執行對照，防止A、C資訊互流
3.  $A \rightarrow B \leftarrow C$ ：原本A、C互相獨立；以B執行對照，在自圓其說效應下，A、C將變成負相關

**注意！**

統計學程序常把每個可測量項目都拿來執行對照，是錯誤的

# 後門準則

## Back-door criterion

明確界定因果圖中哪些變項為去干擾因子，若能取得這些變項的資料，就可依這些變項進行調整，不須實際介入就可預測介入結果

同時符合以下條件，並阻斷所有後門路徑，就可去除干擾Z

- (1) Z不是X的後代
- (2) Z位於後門路徑上

# 後門路徑

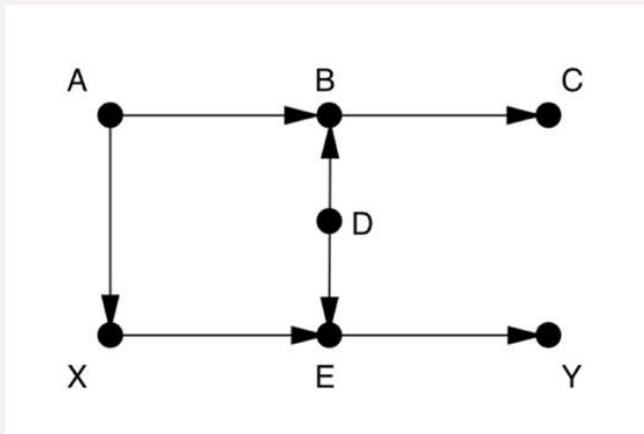
1. 從X指向Y
2. 以指向X的箭頭為起點的路徑

檢驗該變項是否為干擾因子的條件：

須同時滿足以下兩項才為干擾因子，須執行對照

1. 該變項符合干擾的流行病學經典定義
2. 該變項通過後門準則

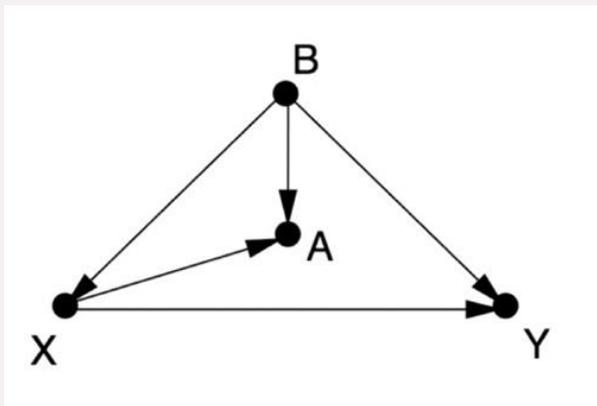
## 範例說明一.



存在後門路徑 $X \leftarrow A \rightarrow B \leftarrow D \rightarrow E \rightarrow Y$ ，  
但此路徑已被衝突B阻斷，因此不用  
執行任何對照

**注意！若以B執行對照反而將開  
啟後門路徑，造成干擾**

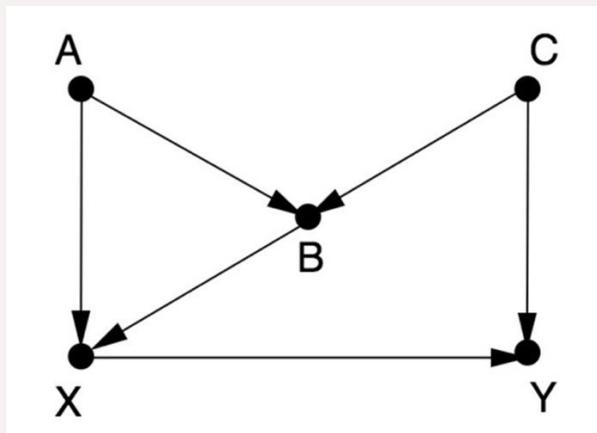
## 範例說明二.



存在後門路徑 $X \leftarrow B \rightarrow Y$ ，需以B執行條件化將其阻斷

**注意！**  
若B無法觀察，則需進行RCT

## 範例說明三.



存在兩條後門路徑：

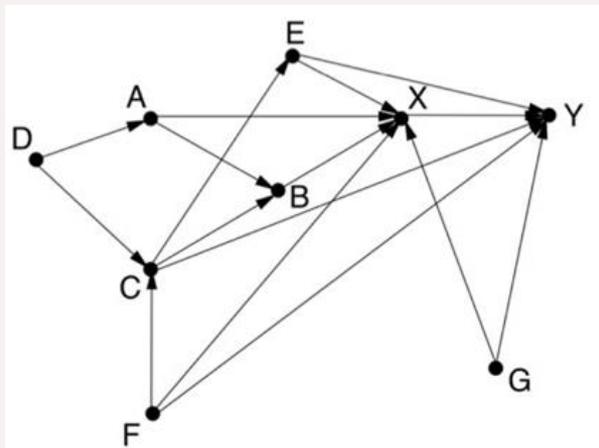
1.  $X \leftarrow A \rightarrow B \leftarrow C \rightarrow Y$

2.  $X \leftarrow B \leftarrow C \rightarrow Y$

若以B執行對照，可封閉2.但會開啟1.；  
想要封閉1.須同時再以A或C執行對照

因此我們最後決定只以C執行對照，  
如此可封閉2.且不影響其他路徑

## 範例說明四.



存在五條後門路徑：

1.  $X \leftarrow E \rightarrow Y$
2.  $X \leftarrow A \rightarrow B \leftarrow C \rightarrow Y$
3.  $X \leftarrow B \rightarrow C \rightarrow Y$
4.  $X \leftarrow F \rightarrow Y$
5.  $X \leftarrow G \rightarrow Y$

由於C為不可觀察的氣喘因素，因此無法項前例一樣直接以C執行對照，必須同時以A和B執行對照

**最後我們以A, B, E, F, G進行對照**

# Takeaway

干擾偏差

隨機對照試驗：消除干擾

do運算子

以貝氏網路三種連結執行對照

後門路徑

後門準則

Thank You :)