【統計解析入門①】

獨協医科大学
先端医科学統合研究施設　研究連携・支援センター
西連地　利己

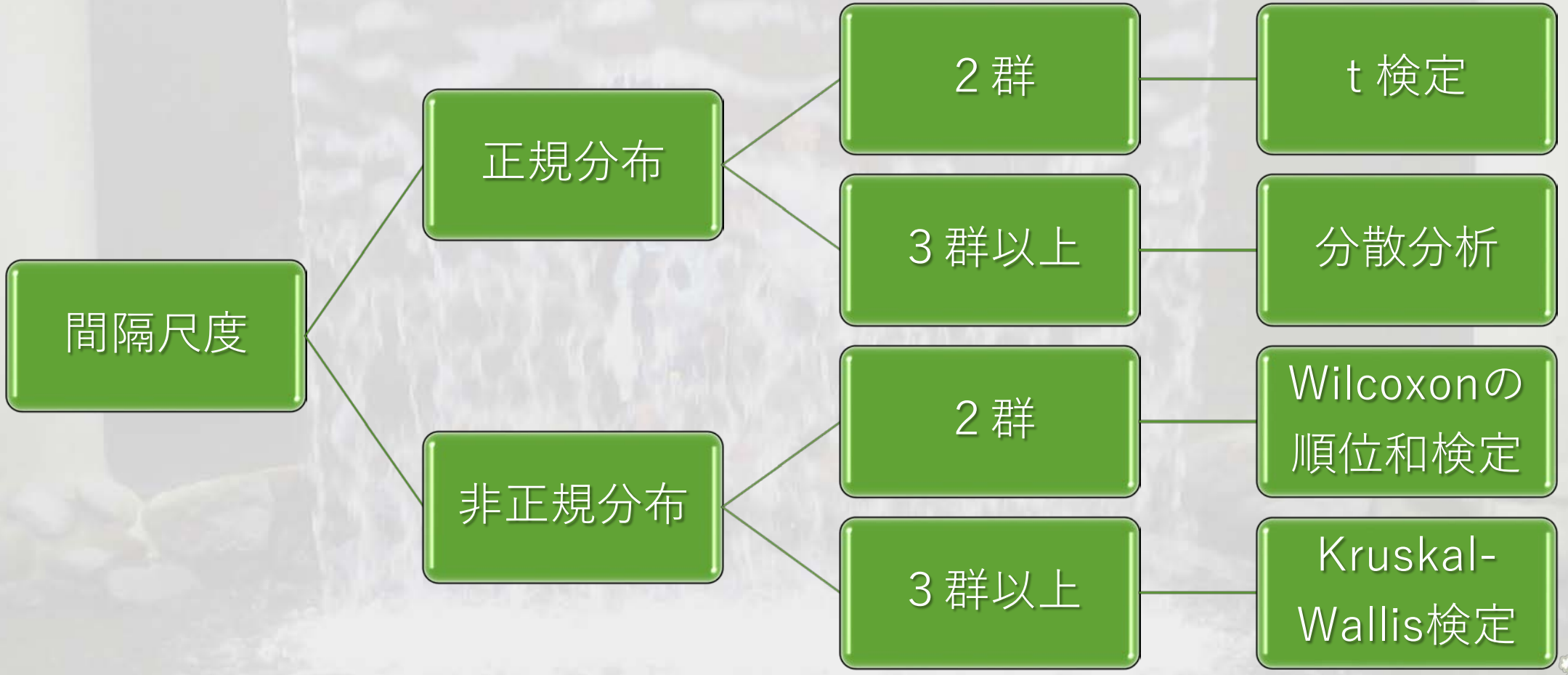
解析手法はデータに応じて選ぶ

- 間隔尺度ですか？
  - 正規分布していますか？

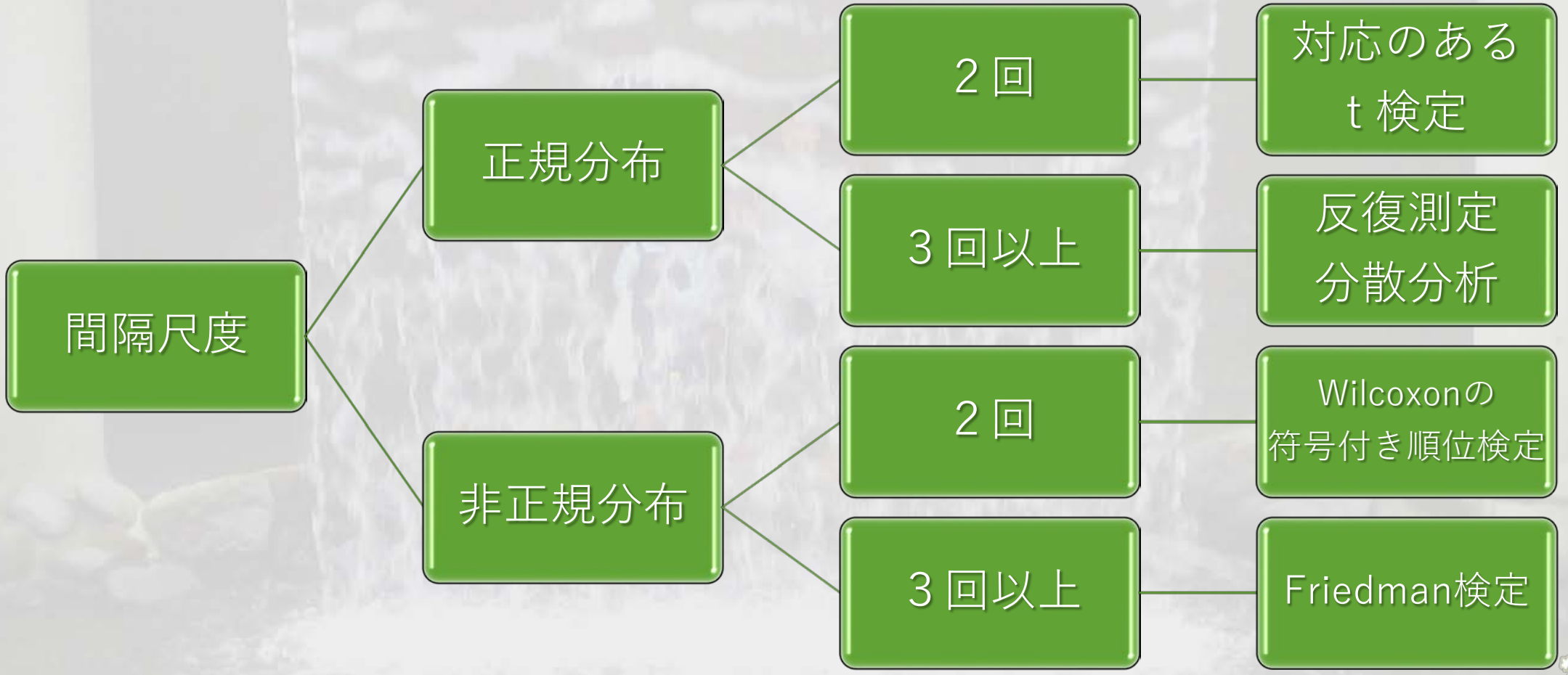
- 名義尺度ですか？

単変量解析

間隔尺度の群間差の有無



間隔尺度の変化の有無


間隔尺度
正規分布
非正規分布
2回
3回以上
2回
3回以上
対応のある
t検定
反復測定
分散分析
Wilcoxonの
符号付き順位検定
Friedman検定

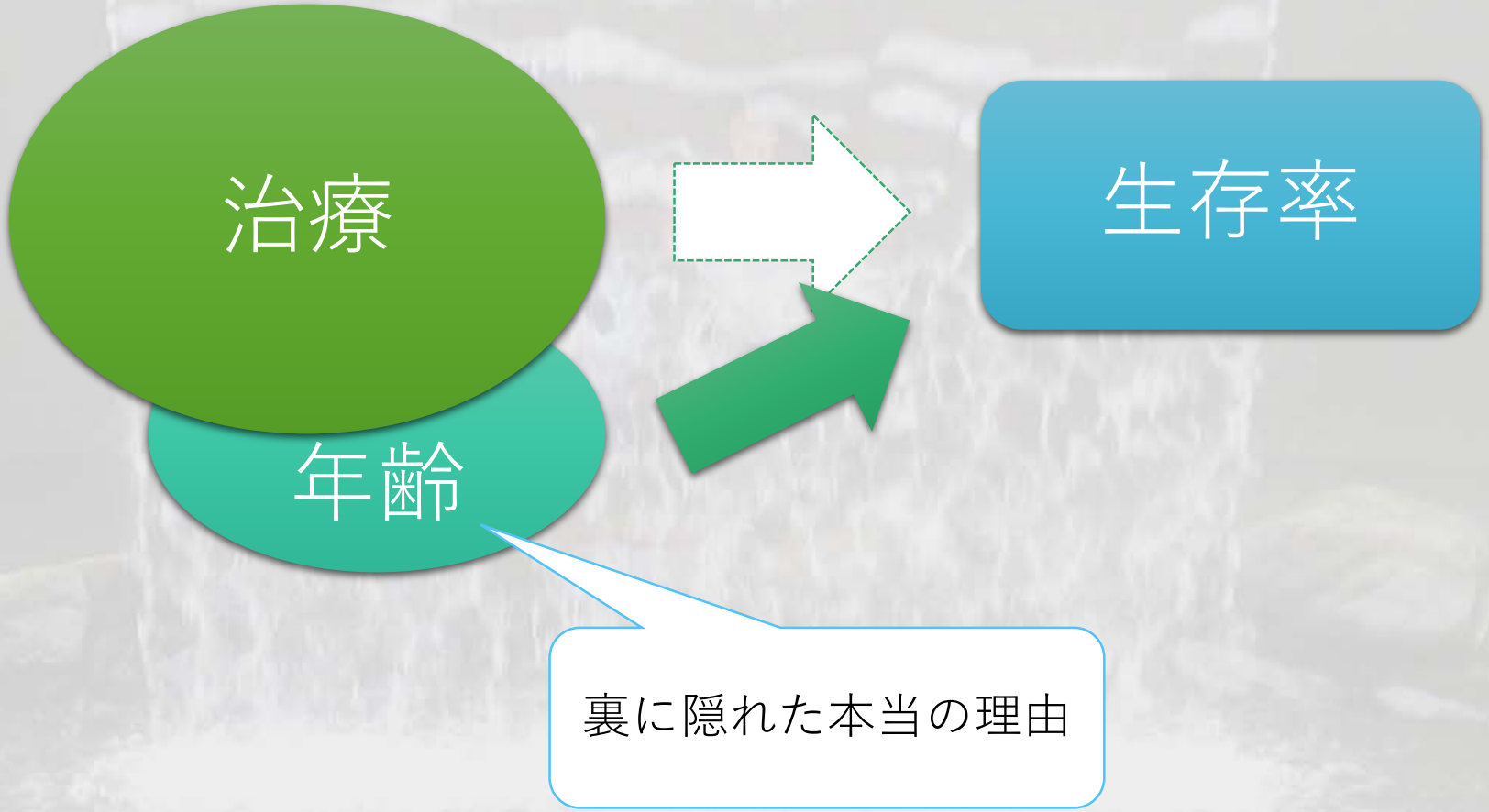
間隔尺度と間隔尺度の関連

- どちらも正規分布している場合
  - Pearsonの積率相関係数
  - 単回帰分析
- どちらかが正規分布していない場合
  - Spearmanの順位相関係数
- 従属変数が正規分布している場合
  - 単回帰分析

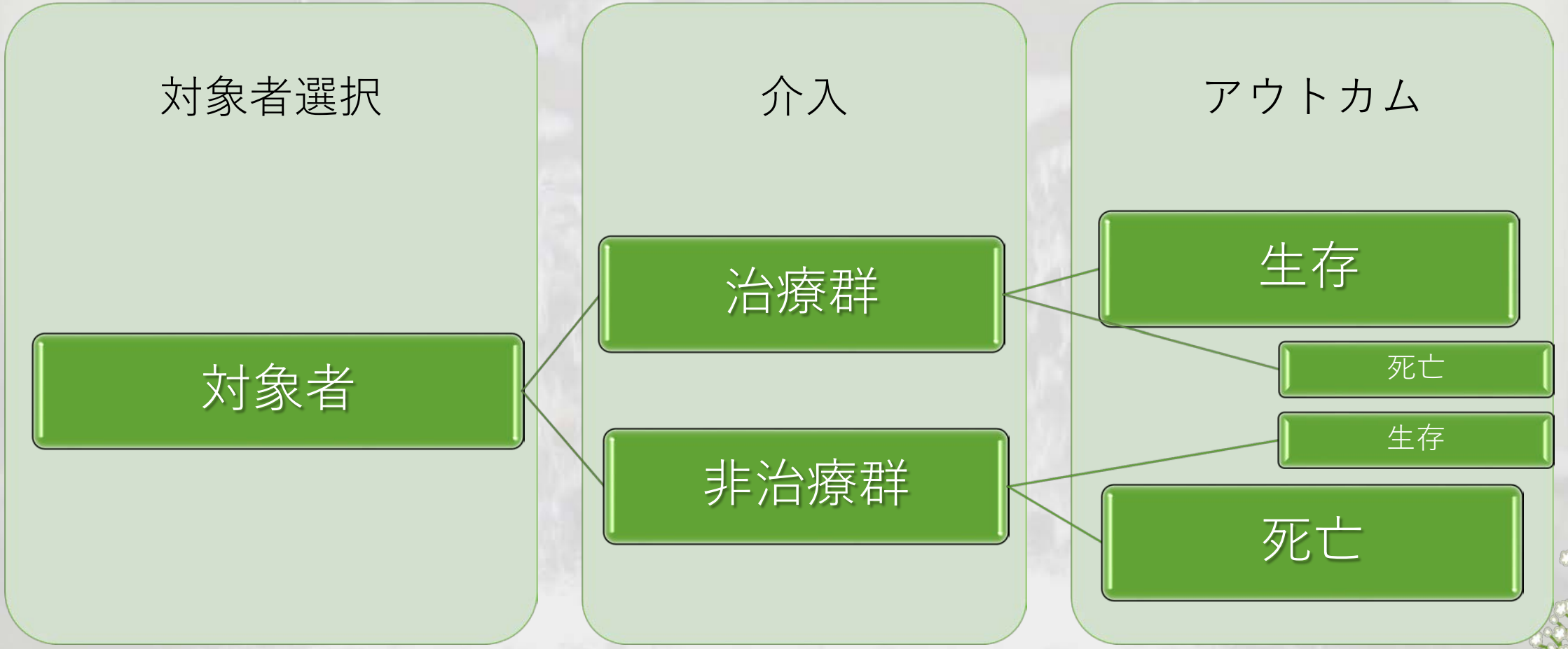
名義尺度（割合）の群間比較

カイ２乗検定

交絡 confounding



RCT (Randomized Controlled Trial)

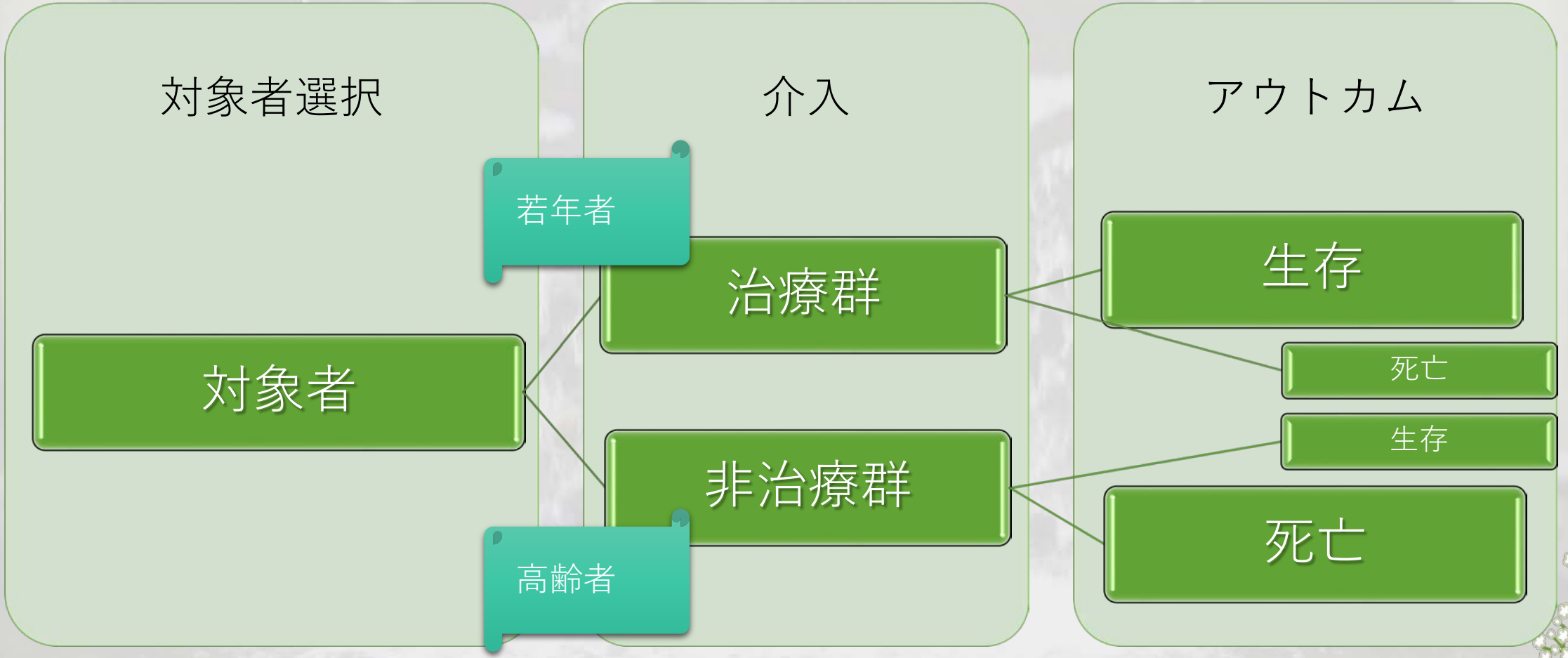


ランダム割付のメリット

- 治療以外の背景因子が、治療群と非治療群で同じになる。
  （確率的な話ですが・・・）
    - 合併症、本人の意思、性、年齢、収入、・・・・

単変量解析で十分

もし、群の年齢が異なっていたら・・・



交絡因子

- ベースライン時に群分け（割り付け）した際に
  - 群間で異なり（即ち群分け変数と相関があり）、
  - かつ、アウトカムに関連する因子
    - 平均年齢、性比、喫煙率、病態、・・・・・

アウトカムに影響を与えたのは、
治療ではなく、これらの因子であることが、
否定できない！！

交絡のチェック①

項目	治療A群	治療B群	p
年齢（歳）	41.2 ± 3.3	41.2 ± 3.3	0.999
男性（%）	20.3	20.3	0.999
喫煙率	14.2	14.2	0.999
重症度高値（%）	45.3	45.3	0.999
・・・	・・・	・・・	・・・
・・・			・
・・・		・	・・・

群間で全く差のないこれらの因子は
アウトカムの差を生まない　はず

交絡のチェック②

項目	治療A群	治療B群	p
年齢（歳）	41.2 ± 3.3	34.1 ± 3.2	0.043
男性（%）	20.3	20.4	0.430
喫煙率	14.2	15.5	0.124
重症度高値（%）	45.3	74.2	0.012
・・・	・・・		・・・
・・・			・・
・・・			・・・

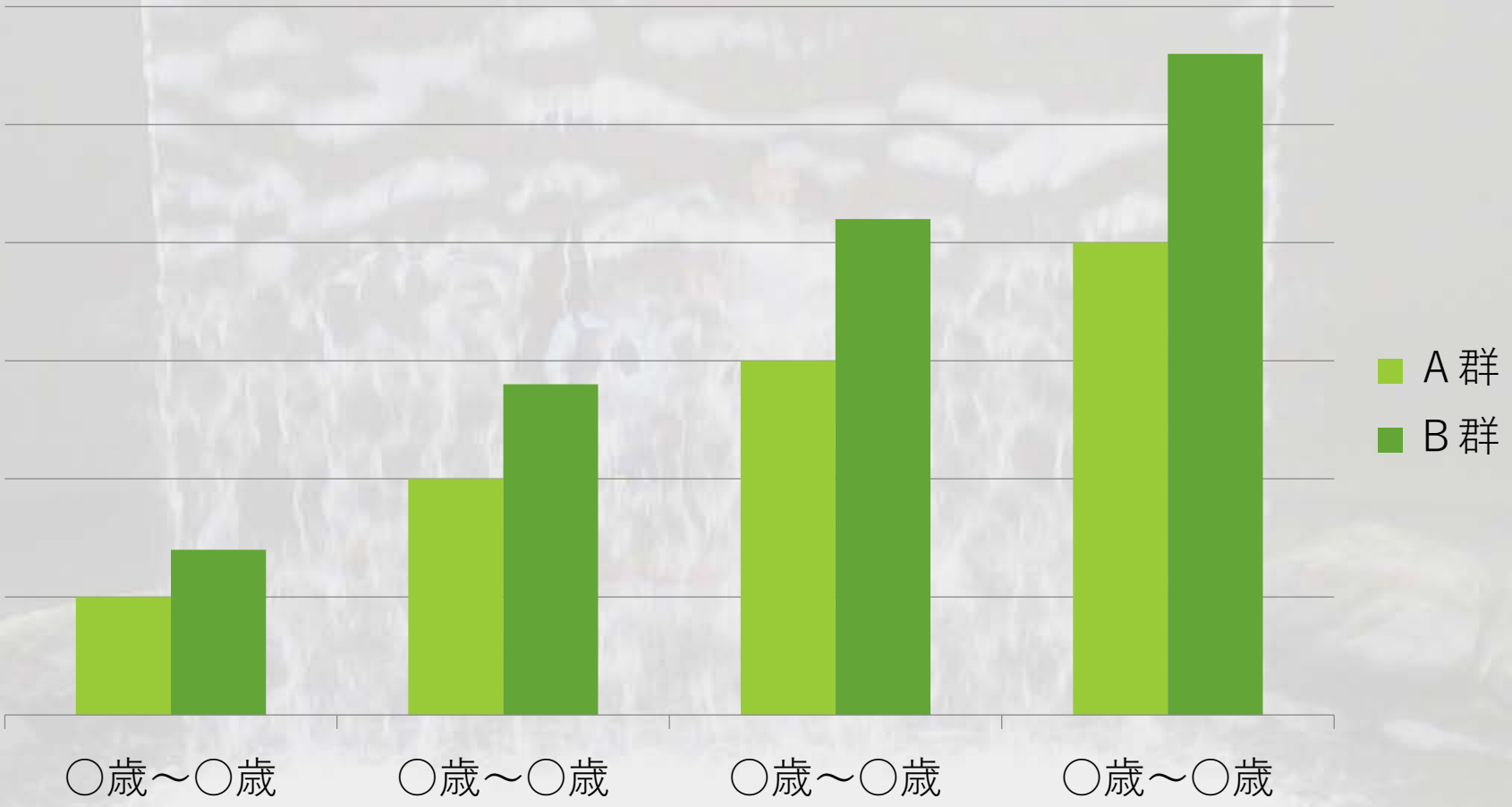
有意差のある因子は
アウトカムの差を生む可能性あり
※有意差が無くても交絡になる場合あり

交絡の制御方法

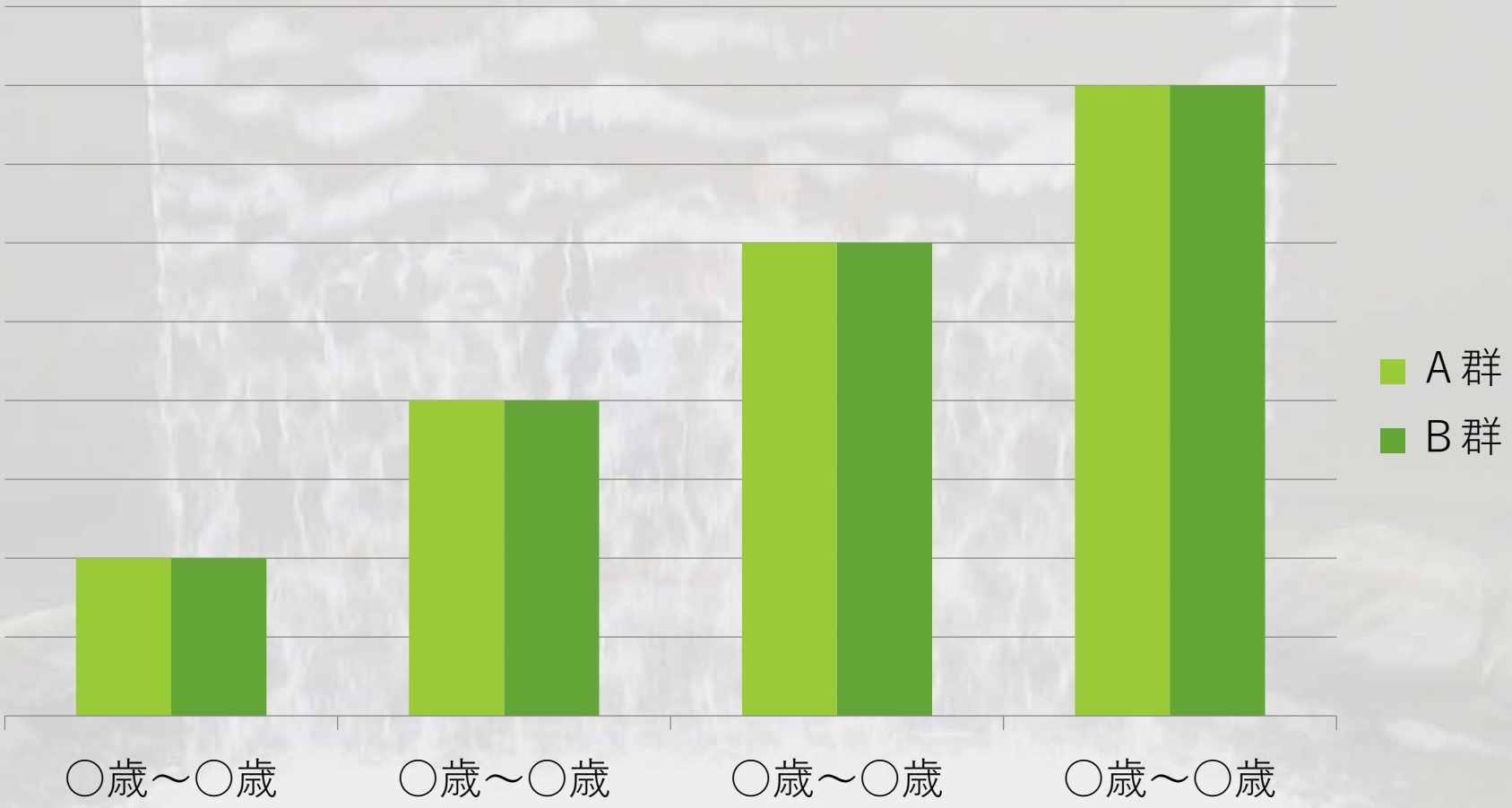
- 例：平均年齢の違いを除去して、A群とB群との比較をしたい

- 層別化
- マッチング
- 多変量解析

例：年齢による層別化



例：年齢による層別化

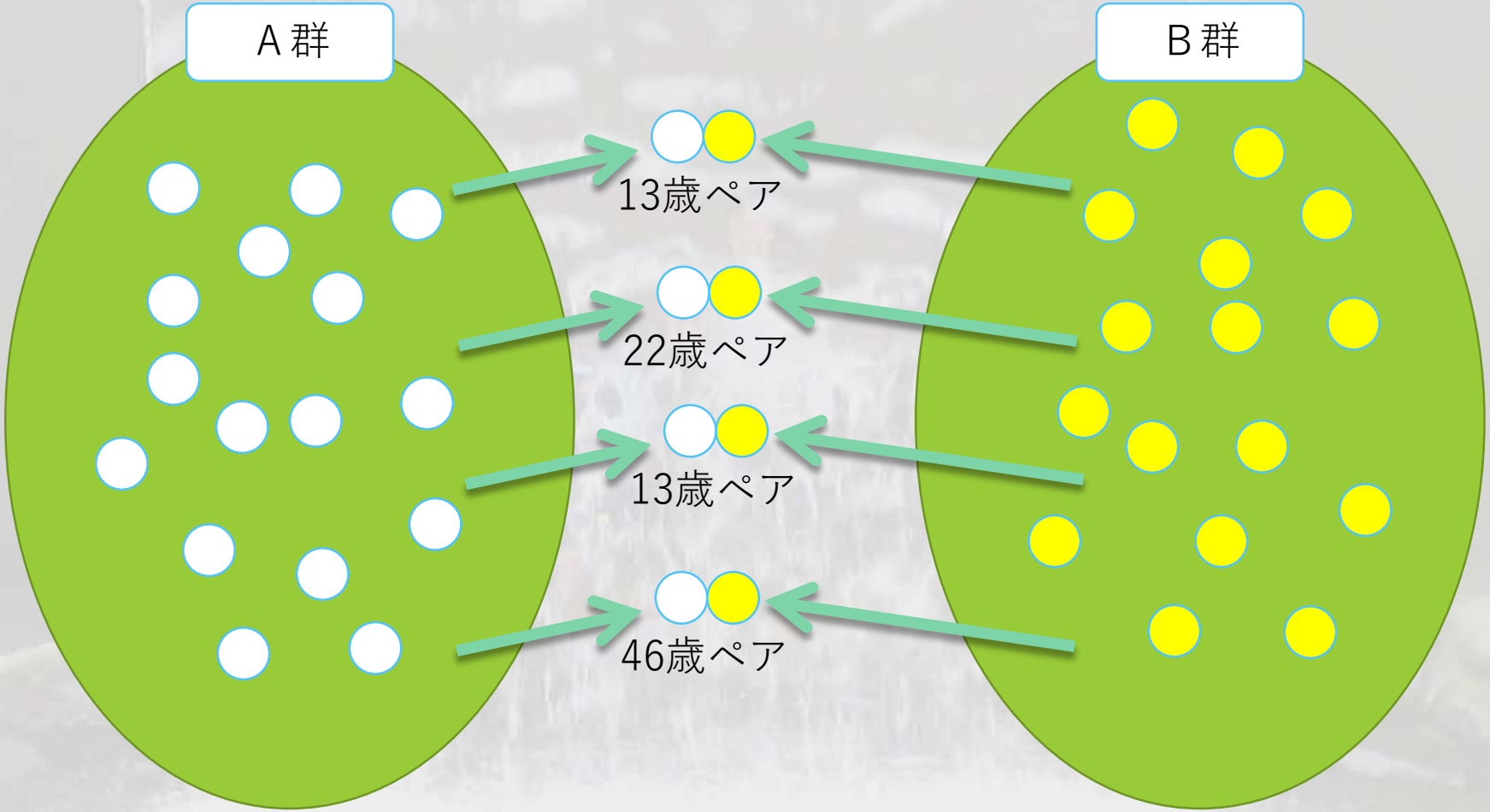


層別解析の問題点

- 層が多くなると大きなサンプルサイズが必要
  - 男女と年齢３階級と治療前病態4階級
    - 2×3×4= 24の層
    - それぞれに100人必要なら、合計2400人必要

- アウトカムの有意差が有る層と無い層があったら・・・
  - 解釈が難しくなる場合も・・・

例：年齢でマッチング



マッチングの問題点

- マッチングできないデータは除外される

- たくさんの項目でマッチングしようとするとマッチング相手がいなくなる
  - 男性で32歳で１日に30本の喫煙者でステージIIIで〇〇病の既往歴があって・・・・・

そこで多変量解析

- （完璧ではないが）交絡因子を調整可能
- 既存データを有効活用

次回に続く