

疫学で用いられる統計的手法（生物統計学）、バイアス

ーデータそのものや、得られた結果を正しく理解するためにー

精度、再現性（Precision、Reliability） → 誤差（Error、Random error）
統計的推測の対象

正確性、妥当性（Accuracy、Validity） → バイアス（Bias、Systematic error）

標本→母集団（一般化）

（標本から）得られた結果は絶対的なものではない

尺度の信頼区間：どの程度確からしいか

95%信頼区間などが用いられる（既出）

研究結果を歪めるもの：バイアス（観察研究において、制御不能・不足から生じる）

1. 研究計画、対象者選択時

admission rate (Berkson), assumption, autopsy series, centripetal, design, diagnostic access, diagnostic purity, digit preference, lead time, length, membership, migrator, missing clinical data, non-contemporaneous control, popularity, prevalence-incidence (Neyman), procedure selection, referral filter, sampling, selection, starting time, susceptibility, unacceptable disease, volunteer, wrong sample size

2. 実施時

apprehension, ascertainment, attention, bogus control, compliance, contamination, detection, diagnostic suspicion, diagnostic vogue, end-digit preference, expectation, exposure suspicion, family information, insensitive measure, instrumental error, interviewers, missing values, obsequiousness, observer, performance, previous opinion, recall, response, substitution game, therapeutic personality, transfer, unacceptability, underlying cause, unmasking (detection signal), withdrawal

3. 解析時

data dredging bias, estimator, in handling outliers, post-hoc significance, repeated peeks, scale degradation, tidying-up

4. その他

the all's well literature, cognitive dissonance, confounding, correlation, ecological, hot stuff, interpretation, magnitude, mimicry, mistaken identity, one-sided reference, positive results, publication, reporting, rhetoric, significance, under-exhaustion, wish

- ・対象者選択時のバイアス：選択バイアス（Selection bias）
- ・調査実施時のバイアス：情報バイアス（Information bias）→誤分類（Misclassification）
- ・交絡（Confounding）

選択バイアス

偏った集団を対象者として選択してしまう。

→ 一般化不可能（ex. Healthy Worker Effect）

誤分類

曝露評価や疾病評価をする際に、系統的に誤った評価（分類）をしてしまう

Differential misclassification, Nondifferential misclassification

コホート研究での曝露分類をする際に生じる誤分類は、疾病発生の有無とは無関係に生じると考えるのが一般的であろうし (nondifferential misclassification)、一方ケースやコントロールであることの情報が **masking** されていない状況では、インタビューによって曝露分類を行おうとする場合に、ケース側により詳細なインタビューを行い、コントロールとの間で曝露把握の正確性に差をもたらす可能性がある (differential misclassification)。

曝露側あるいは疾病側に **nondifferential misclassification** が生じた場合は、リスク比やオッズ比の値を、真の値から「影響なし」の方向に歪め (比の尺度であれば1に、差の尺度であれば0に近づける)、**differential misclassification** の場合は歪める方向が一定しない

感度、特異度がわかれば調整は可能 (but わからないからバイアスになる)

cf. Measurement error

交絡および交絡要因 (Confounding factorまたはConfounder)

- ・ 非曝露グループが「曝露を受けなかった場合の曝露グループ」と同じにならない場合、交絡が存在する。 → 比較グループ間の構成にバランスがとれていない
- ・ 交絡要因の必要条件：
 1. 交絡要因は疾病のリスクファクター
 2. 交絡要因は曝露変数と関連している
 3. 交絡要因は、曝露と結果の因果連鎖の間にある変数ではない
- ・ 交絡の制御：
 1. 制限 (限定)
 2. マッチング (効率化)
 3. 層別解析 (重み付き平均)
 4. 多変量モデル (Black Box的解析)

cf. 効果修飾要因 → 交互作用を引き起こす
モデル等による修正は不可能
層ごとに解析する必要あり

層別解析 (主に年齢階級や性別による調整)

1. 直接法による調整 → (年齢) 調整死亡率 ((Age) Adjusted mortality rate)
2. 間接法による調整 → 標準化死亡比 (Standardized Mortality Ratio, SMR)
3. マンテル・ヘンツェル (Mantel-Haenszel) 推定 → 要約率比・リスク比・オッズ比

SMRとMantel-Haenszel推定値

1. 以下のいずれかの場合、両者は類似の値を示す。
 - ・ 各層で観察された比に大きな違いがない場合
 - ・ 曝露群、非曝露群でも、観察された人-時間の合計が各層で大きな違いがない場合
 - ・ 各層において、非曝露群の人-時間が曝露群のそれに比べてきわめて大きい場合
2. 各層の値がほぼ均一の場合Mantel-Haenszel推定値を、各層の値が異なる場合SMRを用いる。
注意：層別解析 (および次の多変量モデル) は各層の影響の尺度がほぼ一定の場合に使用すべきもので、層ごとの値が大きく異なる場合 (影響のあるなしが逆になっている場合など) は、要約・調整してはいけない

多変量モデル（疫学研究でよく用いられる手法）

1. ロジスティック回帰分析
二値データ、調整オッズ比
2. 比例ハザードモデル
生存時間データ、時間によらずハザード比一定の仮定、調整率比
3. ポアソン回帰分析
頻度データ（急性死亡データ）、調整率比

モデル作成・解釈上の注意事項（だぶる点もあり）

1. 使用する変数
2. 変数選択（1とは異なる）
3. 多重共線性（主成分回帰などにより回避→グラフィカルモデリング）
4. モデルの妥当性
5. モデルの限界の把握

通常の統計モデルはデータに独立性を仮定しているが、現実的には相関性を持つデータを扱うことの方が多。相関性を認めたモデルも最近ではよく利用される。

→相関性を認めた解析の例：GEE（Generalized Estimation Equations、一般化推定方程式）