「疫学・生物統計学」講義概要

- テキスト p.29 〜
- 疫学とは何か，成り立ちとリスク要因について説明した後，因果推論について説明する。
- 参考文献
  - 佐藤俊哉・松山裕「疫学・臨床研究における因果推論」 In:『統計科学のフロンティア５　多変量解析の展開　隠れた構造と因果を推理する』(岩波書店，2002）
  - Rothman KJ（矢野・橋本訳）「ロスマンの疫学」(篠原出版新社，2004）
- 生物統計学のさまざまな手法について説明する。

疫学（epidemiology）とは？

- 国際疫学会「疫学辞典」第4版の定義『特定された集団における健康に関連した状態あるいは事象の分布と決定因子の研究，及び，この研究の健康問題の制圧への応用』
- 日本疫学会編「疫学」における柳川の定義『明確に規定された人間集団の中で出現する健康関連のいろいろな事象の頻度と分布およびそれらに影響を与える要因を明らかにして，健康関連の諸問題に対する有効な対策樹立に役立てるための科学』
- 『明確に特定された人間集団の中で出現する健康に関する様々な事象の頻度及び分布並びにそれらに影響を与える要因を明らかにする科学研究をいう』(2002 年 6 月 17 日，文部科学省・厚生労働省「疫学研究に関する倫理指針」における「疫学研究」の用語定義）

有名な疫学研究

- Snow J (1813-1858) のコレラの研究：当時コレラ菌は未発見であり，ロンドンでは頻繁にコレラが流行していた。1854 年にロンドンのブロードストリートで大流行があった際に，Snow はコレラ死亡者の発生地図，死亡の日別分布表を作成し，詳細な症例検討を加えて，流行の原因となった共同井戸のコレラ菌汚染を突き止めた。記述疫学研究。
- Panum PL (1820-1885) の麻疹の研究：デンマーク領ファロー島はそれまで 65 年間なかった麻疹が 1846 年に大流行したので，Panum は数千人の患者を診察して，大流行の原因，感受性と終生免疫，接触から発症までの期間，感染可能期間などを明らかにした。
- Gregg NM：1941 年にシドニーなどで多発した先天性白内障の観察から，その前年に起こった風疹流行と母親の当該児妊娠初期が一致することを発見した。
- 高木兼寛（1849-1915) の脚気の研究：1884 年に脚気の原因が食事の欠陥にあると見当をつけ，脚気が多発していた海軍の軍艦乗組員の遠洋航海の際の食事を変えて，大麦，大豆，牛肉を増やすことによって，炭素に対する窒素割合を増やしたら，脚気が減った。真の原因がビタミン B1 不足であることまではわからなかったが，この介入研究によって，食事に原因があることは明らかになった。

疫学研究のフレームワーク

- ”5-W-Bridge”：疾病について，いつ，どこで，誰が，どんな病気に，何故罹ったかを明らかにできれば，原因も突き止められるということ。そのために必要な手順として，
  - 疾病分類の明確化：共通の分類基準が必要。
  - 調査対象または調査資料の選択：どういうデータを使うか？
  - 調査すべき疾病量の把握：集団における疾病罹患状況を示す指標の性質を把握する。
  - 調査方法の選択：記述疫学か分析疫学か介入研究か？
  - 調査の実施と結果の分析：基本的に統計学を利用する。
  - 結果の解釈と評価：因果関係の判断は難しい。
- 以下，これらの手順について個別に説明する

疾病分類

- 疾病分類＝ある一定の基準により疾病を分類する体系。疾病の単位を明確にし，異なる調査結果を比較することを可能にする。すべての疾病について漏れが無く重複もないのが理想。
- 国際疾病分類（International Classification of Diseases = ICD）：WHO の前身である国際会議の協議により 1900 年に制定され，約 10 年毎に改定され，1995　年から第 10 回修正国際疾病分類（ICD-10）が使われている。ICD-9 では4桁（3桁＋小数点＋1桁）の数字で分類，ICD-10 では最初の文字を U 以外のアルファベットにしたので最大 24999 種類まで分類できる。実際の項目数も約 7000 から約 14000 に倍増。既存資料に基づいて死亡率や罹患率の長期的な変化を調べるときは，分類に使われている ICD の回の違いによるカテゴリの違いの影響を受けないように工夫する。この細かい分類を基本分類といい，それに対して，とくに死因に対しては，個々の疾病を約 130 項目にまとめた死因分類というものを用いるのが普通である（ICD-9 では死因簡単分類と呼ばれた）。国連や WHO の統計資料は基本分類でなくて死因分類や死因簡単分類でまとめられている
- 死亡診断書：人口動態統計データを利用する上では，その元になる死亡診断書の内容を把握すべきである（テキスト p.32，図 3-1 を参照）。人口動態統計に死因として記載されるのは，周産期死亡を除き原死因であることに注意（それでは不都合な場合はある一定の準則により修正される）。

使うデータ

- 調査をするにせよ，既存資料を使うにせよ，まずリスク曝露人口（population at risk）の特定が必要。リスク曝露人口は疾病の程度を示すための分母となる。例えば，子宮ガンでは女性全員。国民全体を対象とする場合は，国勢調査による日本人口を使うのが普通。厳密に考えれば7月１日時点の人口を年央人口として使いたいところだが，国勢調査では 10 月１日時点の推計人口が提示されるので，それを年央人口として使うのが普通である。
- 厳密なリスク曝露人口の把握が困難な場合は，別の測定しやすい値で代用することもある。
- 既存資料を使う場合は，どのように実施された調査の結果を，どのような指標を使ってまとめた資料なのかという点を吟味して扱う必要がある。

標本抽出法とは？

- 疫学調査で得られた結果を適用したい集団（介入研究でいう参照集団，統計的には母集団）の全数を調査する悉皆調査は，費用や時間などの制約，あるいはその必要がない，回収率が低くてバイアスが大きい等の理由で実施されないことが多い。
- その代わりに，集団全体を代表する適当なサイズの標本をうまく選んでやればよいことになる。
- 如何にうまく集団全体を代表するような標本を選ぶかという目的で考案されたさまざまな方法を総称して標本抽出法と呼ぶ。
- 単純無作為抽出，層化無作為抽出，集束抽出，確率比例抽出など。

有病割合（prevalence）

- 「ある集団のある一時点で疾病ありの人数」を「ある集団の調査対象人数」で割った値
- 意味：急性感染症で有病割合が高いなら患者が次々に発生していることを意味するが，慢性疾患の場合はそうとは限らない。
- 応用：行政施策として必要な医療資源や社会福祉資源の算定に役立つ
- 例：高血圧や高コレステロール血症の有病割合が高い。

罹患率（incidence rate）

- 発生率。個々の観察人年の総和で発生数を割った値。次元は1／年。
- “A Dictionary of Epidemiology, 4th Ed.” に明記されているように，incidence は発生数。
- 感受性の人の中で新たに罹患する人が分子。再発を含む場合はそう明記する必要がある。
- 意味：瞬時における病気へのかかりやすさ。つまり疾病罹患の危険度（リスク）を示す。
- 疾病発生状況と有病期間が安定していれば，平均有病期間＝有病割合／罹患率

年齢による標準化の方法

- 死亡率の場合なら，直接法年齢調整死亡率と間接法年齢調整死亡率となる
- 直接法は，対象集団の年齢構成が基準集団と同じだった場合に対象集団の年齢別死亡率に従って死亡が起こったら全体としての死亡率はどうなるかと考える思想。情報としては対象集団の年齢別死亡率が必要
- 間接法は，対象集団が基準集団の年齢別死亡率に従って死んだ場合に期待される死亡数で実際の対象集団の死亡数を割って標準化死亡比（SMR）を出し，それに基準集団の死亡率を掛けて得る。対象集団についての情報としては，年齢別人口と総死亡数がわかっていれば十分

相対危険（Relative Risk）

- 曝露群のリスクの対照群のリスクに対する比
- リスクとして累積罹患率をとると，累積罹患率比（cumulative incidence rate ratio）またはリスク比（risk ratio）
- リスクとして罹患率をとると，罹患率比（incidence rate ratio）
- リスクとして死亡率をとると，死亡率比（mortality rate ratio）
- 罹患率比と死亡率比を合わせて率比（rate ratio）という

オッズ比（Odds Ratio）

- オッズ（ある事象が起きる確率の起きない確率に対する比）の比
- 2 種類のオッズ比（コホート研究における累積罹患率のオッズ比と患者対照研究における曝露率のオッズ比）は数値としては一致する。
- オッズ比は（稀な疾病の場合）率比の近似値として価値がある
- 交絡因子を調整してオッズ比を出すにはロジスティック回帰分析など

	病気あり	病気なし
曝露あり	a 人	b 人
曝露なし	c 人	d 人

左の表のような観察結果があるとき，コホート研究における疾病オッズ比 (disease odds ratio) は $(a/b)/(c/d)=(ad)/(bc)$ となり，患者対照研究における曝露オッズ比 (exposure odds ratio) は $(a/c)/(b/d)=(ad)/(bc)$ となって一致する。断面研究における有病割合オッズ比も一致する。

誤差

- 誤差＝真値との差
- 標本誤差＝標本抽出の偶然変動に伴う母集団の真値との差（標本サイズが大きいほど小さい）
- 非標本誤差＝標本誤差以外の誤差（例えば不適切な標本抽出による誤差）
- ランダムな誤差（偶然誤差）
  - ランダムな誤差が小さい＝精度（precision）が高い
- 系統的な誤差＝バイアス
  - 研究デザイン，データ収集，分析，レビュー，出版など，研究のさまざまな段階で起こりうる。
  - 系統的な誤差が小さい＝正確さ（accuracy）が高い

バイアスの制御

- 無作為化（randomization）：ランダムに群分けをすることで介入以外の条件を確率的に均質化
- マッチング（matching）：交絡因子の条件が似るように対照群を選ぶ
- 層化（stratification）：交絡因子のカテゴリ別に解析することで交絡因子の影響を除く
- 標準化（standardization）：基準集団を決めて交絡因子のカテゴリ別割合を調整することで交絡因子の影響を取り除く
- 限定：交絡因子の1つのカテゴリに属する者だけを対象として分析することで，交絡因子の影響を取り除く

因果関係

- 疫学の目的は，集団の健康を増進し守り保つことであり，そのために健康に影響する要因を明らかにするのだから，要因→影響の関係をはっきりさせることが大事である。
- しかし，相対危険と過剰危険のところで触れたように，そこをはっきりさせることは難しい。
- 要因と影響の間に何らかの関連があることは，数学的には相関関係で表される。が，因果関係があるかどうかは別物である。
- もっとも強い因果関係は，生物学的に，要因が影響を起こすメカニズムが明らかであり，それが常に成り立つ場合にいえる（生物学的因果関係）。

個人レベルでの因果関係は立証可能か？

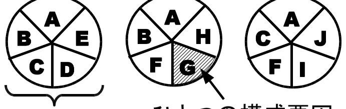
- 例えば，香港に旅行することがSARS罹患の要因となるかどうかを知りたいとする
- Aさんという個人に着目すると，Aさんは香港に旅行するかしないかどちらかの行動しかとりえないし，SARSに罹るか罹らないかのどちらかの結果しかありえない
- つまり，Aさんが香港に旅行してSARSに罹ったという事実があるときに，Aさんが香港に旅行しなかったからSARSに罹らなかったという事実はありえないので，個人レベルでの因果関係は立証不可能である
- こういう考え方を反事実（counterfactual）モデルという。
- すると，目の前の患者さんの病気について，因果関係を明らかにすることはできないのか？

集団レベルでの因果関係

- 喫煙していて肺がんにかかったAさんが喫煙しなかった場合の観察事実はありえないが，喫煙という曝露条件をもつ集団Aに対して，**喫煙以外の条件がほとんど同じ**集団Bを設定して，集団Aと集団Bの間で肺がんの発生率を比較することはできる。
- 集団レベルの因果関係がいえれば，個人でもその可能性が高いだろうと推論できる
- 実は，この集団AとBが，喫煙という要因が肺がんという疾病に影響する関係における，曝露群と対照群である。
- 「喫煙以外の条件がほとんど同じ」になるように対照群を選ぶ操作をマッチングと呼ぶ。
- ただし，患者対照研究ではマッチングに注意が必要

因果パイモデル (Rothman, 2002 など)

- 複数の因子を含む因果関係では，その因子の組み合わせは一通りとは限らない。
- それらの因子が揃えば必ず疾病が起こるという条件の組を十分要因群（sufficient causes）という。十分要因群を構成する個々の要因を構成要因（component causes）という。
- 十分要因群の組を円グラフの形で表したものを因果パイモデルと呼ぶ（下図は，ある疾病を引き起こす3種の十分要因群；出典は Rothman, 2002）



ひとつの因果メカニズム　ひとつの構成要因