

疫学・生物統計学資料, Rev.9.1

中澤 港

April 12, 2024

初版：2014年11月17日
第2版：2015年1月25日
第3版（体裁変更）：2016年10月2日
第4版（倫理指針関係の記述更新）：2017年10月9日
第4.1版（索引付け途中まで）：2018年2月15日
第5版（索引付け継続、ICD-11の記述追加）：2019年10月4日
第6版（スクリーニング大きく加筆修正）：2020年8月25日
第7版（致命割合について加筆）：2021年2月23日
第8版（文献追加）：2021年10月9日
第8.1版（微修正）：2022年11月9日
第9版（諸情報更新）：2023年5月1日
第9.1版（参考文献追加、ICD-11の記述追加）：2024年4月12日

Contents

1	概論	7
1.1	参考文献	7
1.2	疫学 (epidemiology) の定義と目的	8
1.3	疫学研究のフレームワーク	8
1.4	疾病分類 (classification of diseases)	9
1.5	データソース	10
2	疫学研究のデザイン	11
2.1	観察的疫学研究のいろいろ	11
2.2	記述疫学	12
2.3	分析疫学	12
2.3.1	自然実験	12
2.3.2	生態学的研究 (地域相関研究)	12
2.3.3	横断的研究	14
2.3.4	症例対照研究	14
2.3.5	コホート研究	15
2.3.6	より工夫した観察研究	15
2.4	介入研究	16
3	標本抽出法	17
3.1	単純無作為抽出法	17
3.2	層別抽出法 (層化抽出法)	18
3.3	集落抽出法	18
3.4	確率比例抽出法と副次抽出法	18
3.5	標本抽出法の選択	19
3.6	サンプルサイズ	19
4	疾病量の把握	21
4.1	疾病量の指標	21
4.1.1	有病割合 (prevalence) とオッズ (odds)	22
4.1.2	罹患率 (incidence rate)	22
4.1.3	死亡率 (mortality rate)	23
4.1.4	累積罹患率 (cumulative incidence rate=risk)	23
4.1.5	致命割合 (case fatality ratio または case fatality risk; CFR)	23

4.1.6	死因別死亡割合 (proportional mortality ratio; PMR)	24
4.1.7	PMI (proportional mortality indicator) = 50 歳以上死亡 割合	24
4.2	年齢による標準化	24
4.2.1	直接法の考え方	25
4.2.2	間接法の考え方	25
5	危険因子とその影響 (効果) の指標	27
5.1	危険因子とは?	27
5.2	影響 (効果) の指標	27
5.2.1	オッズ比の計算例	28
5.3	寄与割合・人口寄与割合	29
5.4	相対危険と超過危険の関係	29
5.5	生涯リスク	29
6	因果関係	31
6.1	因果関係とは?	31
6.2	第一種の過誤・第二種の過誤	31
6.3	生物学的因果関係を導く指針 (不完全)	32
6.4	個人レベルでの因果関係は立証可能か?	33
6.5	集団レベルでの因果推論のロジック	33
6.6	いろいろな因果関係	34
6.6.1	一要因⇒一症状	34
6.6.2	一要因⇒多症状	34
6.6.3	多要因⇒一疾病	34
6.6.4	多要因⇒多疾病	34
6.7	因果関係の整理～因果パイモデル	34
6.7.1	因果パイモデルの効用	35
7	因果関係を歪めるもの～測り間違い	37
7.1	何が因果関係を歪めるのか?	37
7.2	選択バイアス	38
7.2.1	Neyman's バイアス	38
7.2.2	逸話的情報とサンプリングバイアス	38
7.2.3	調査におけるサンプリングバイアス	39
7.2.4	Berkson バイアス	39
7.2.5	自己選択バイアス	40
7.2.6	健康労働者効果 (Healthy Worker Effect)	40
7.3	情報バイアス	40
7.3.1	追跡の偏り	40
7.3.2	リコールバイアス	41
7.4	交絡 (confounding)	41
7.4.1	交絡要因の3つの必要条件 (例: 肥満の高血圧に影響する という因果関係における年齢の関与)	41
7.4.2	交絡要因を見過ぐすと変な因果推論をしてしまう	41
7.4.3	交絡の制御	42

8 疫学研究の歴史と研究倫理	43
8.1 古典的疫学研究	43
8.2 代表的な疫学研究	44
8.2.1 Frammingham Heart Study (フラミンガム研究)	44
8.2.2 久山町研究	44
8.2.3 INTERSALT 研究	45
8.2.4 代表的な症例対照研究	45
8.3 人を対象とする生命科学・医学系研究に関する倫理指針	45
9 実験計画と生物統計学	47
9.1 実験計画における Fisher の 3 原則	47
9.2 実験計画法の発想	47
9.2.1 何杯のミルクティが必要か?	48
9.3 実験計画におけるサンプルサイズ的设计原則	48
9.4 試験配置法	49
9.5 毒性試験	49
9.6 臨床試験	49
9.6.1 臨床試験の 4 段階	50
9.6.2 臨床試験における倫理的な要求	50
9.6.3 臨床試験の手順	50
9.6.4 ITT の理屈	51
9.6.5 欠損値はどうするか	52
9.6.6 治療効果の判定指標 (その意味)	52
9.6.7 新薬の臨床試験における作業仮説	53
10 スクリーニング	55
10.1 スクリーニング (Screening) とは?	55
10.1.1 スクリーニングの目的	55
10.2 スクリーニングにおけるバイアス	56
10.2.1 自己選択バイアス	56
10.2.2 リードタイムバイアス	56
10.2.3 レングスバイアス	56
10.3 スクリーニング実施の原則	56
10.3.1 現代における新しいスクリーニング基準	57
10.4 スクリーニング方法の評価	59
10.4.1 有効性の指標	59
10.4.2 確定診断に用いる検査方法自体の感度・特異度?	60
10.4.3 信頼性	61
10.5 連続量の測定における ROC 分析	61
10.5.1 ROC 曲線とは	62
10.5.2 ROC 曲線を描く例	62
10.6 母子保健分野の代表的なスクリーニング	63
10.6.1 新生児マススクリーニング検査	63
10.7 成人期以降の代表的なスクリーニング	64
10.7.1 特定健診: 高齢者医療確保法による	64
10.7.2 がん検診: がん対策基本法による	65

10.7.3 その他の検診	65
10.8 スクリーニングの参考文献	66
索引	66

Chapter 1

概論

1.1 参考文献

- Krickberg K, Van Trong P, Hanh PTM (2019) Epidemiology: Key to Public Health, 2nd Ed. Springer.
- Rothman KJ, Greenland S, Lash TL (2012) Modern Epidemiology, 3rd Ed. LWW.
- 柳川洋 (2012) 『疫学ノート—基礎から実践まで—』日本公衆衛生協会
- Rothman KJ (2012) Epidemiology: An Introduction 2nd Ed. Oxford Univ. Press (邦訳『ロスマンの疫学—科学的思考への誘い 第2版』篠原出版新社)
- J.L. フライス (著), KR 研究会 (訳) (2004) 『臨床試験のデザインと解析』アーム
- 厚生労働省「治験」ホームページ¹
- ゲルト・ギーゲレンツァー (2010) 『リスク・リテラシーが身につく統計的思考法』ハヤカワ文庫
- デイヴィッド・ザルツブルグ (2010) 『統計学を拓いた異才たち』日経ビジネス人文庫
- 佐藤俊哉 (2005) 『宇宙怪人しまりす 医療統計を学ぶ』岩波書店
- 佐藤俊哉 (2012) 『宇宙怪人しまりす 医療統計を学ぶ 検定の巻』, 岩波科学ライブラリー 194
- 佐藤俊哉 (2024) 『宇宙怪人しまりす 統計よりも重要なことを学ぶ』, 朝倉書店

¹<http://www.mhlw.go.jp/topics/bukyoku/isei/chiken/index.html>

- Bonita R, Beaglehole R, Kjellström T, WHO (2006) Basic epidemiology. World Health Organization. <https://apps.who.int/iris/handle/10665/43541>²

1.2 疫学 (epidemiology) の定義と目的

疫学とは？

特定された集団^aで健康事象^bの頻度 (frequency) と分布 (distribution)^cを調べ、原因との因果関係 (causal relationship) を探り、病気や死亡を減らすことへの貢献を目指す学問

^aspecified population

^b概ね病気だが、病気に限らない

^c時間分布と空間分布がある

以下のような目的に役立つ。公衆衛生政策立案のために必須のツールといえる。

- 疾病発生要因の追究：リスク因子の特定と評価、因果推論
- 疾病自然史の解明：検診計画や治療効果判定には必須
- 疾病予後要因の解明
- 疾病頻度の将来予測：数学モデルを利用した理論疫学研究
- 疾病対策の企画・評価
- 治療効果の判定
- 健康水準の測定
- 地区診断＝対象集団の居住する地域特性にマッチした施策が必要

1.3 疫学研究のフレームワーク

疫学研究のフレームワーク

疫学研究の手順は，“5W-Bridge” に集約される。いつ (When), どこで (Where), 誰が (Who), どんな病気に (What), 何故 (Why) 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、以下の要素が必要となる。

- 疾病分類の明確化：共通の分類基準が必要。

²木原雅子・木原正博監訳の日本語版が https://apps.who.int/iris/bitstream/handle/10665/43541/9241547073_jpn.pdf からダウンロードできる。

- 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？ 全数調査以外では標本抽出法の選択も重要
- 調査方法の選択：記述疫学か分析疫学か介入研究か？
- 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：因果関係の判断には機械的なマニュアルは存在しない。

以下の節で、それぞれ詳細に記述する。

1.4 疾病分類 (classification of diseases)

ICD-10

第10回修正国際疾病分類。時空が異なっても共通の分類体系に依拠して定義された病気が同定できないと、病気の頻度や分布を調べることができない。そのために提案された、国際的に共通する疾病分類が、ICD (International Classification of Diseases)^a。1900年制定で、現行のものは1995年から使われている第10回修正版のためICD-10と呼ばれる。ICD-11が2022年に発効し、各国で実装のための検討が進んでいる。

^a<http://www.who.int/classifications/icd/en/>

疾病分類について要点は下記。

定義 ある一定の基準により疾病を分類する体系

効用 疾病の単位を明確にし、異なる調査結果を比較可能にする

ICD-10 ICDは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。1990年に改訂された第10回修正国際疾病分類 (ICD-10) は、1995年からずっと日本の公式疾病統計に採用されている。最大24999種類まで可能。実際の基本分類は約14000項目。

ICD-11 2018年6月18日にWHOから公表された（そのときの厚労省からのプレスリリースは<https://www.mhlw.go.jp/stf/houdou/0000211217.html>で読める。英語版は<https://icd.who.int/en>で全文読める）。2019年5月のWHO総会に提出、承認され（<https://www.who.int/news-room/detail/25-05-2019-world-health-assembly-update>）、2022年1月正式発効。日本でも公式疾病統計をICD-11に切り替えるための検討中。参考になりそうな文書やサイトを挙げておく。「ICD-11概要について」³、社会保障審議会（統計分科会疾病、傷害及び死因分類部会）⁴からリンクを辿る

³https://jams.med.or.jp/glossary_committee/doc/2019material_s2.pdf

⁴https://www.mhlw.go.jp/stf/shingi/shingi-hosho_164149.html

と、第9回資料2「ICD-11の公的統計への適用について」⁵や、第10回会議⁶など、参考になりそうな情報が得られる。また、社会保障審議会(統計分科会疾病、傷害及び死因分類専門委員会)⁷からは、第26回参考資料1-1「ICD-11分類の表記に用いる用語について」⁸や、同1-2「ICD-11和訳の取り扱いについて(案)」⁹が参考になる。

死因分類 個々の疾病を約130項目にまとめたもの(ICD-9では「死因簡単分類」)。国連やWHOの統計資料は基本分類でなく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 日本では医師が記載。厚生労働省が「死亡診断書(死体検案書)記入マニュアル」¹⁰を公開している。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。異常死だと死体検案書(医師法20条)。死因分類が変わると死因別死亡統計が変わってしまう(1995年の心疾患激減は「心不全」をできるだけ避けるなどの行政指導の効果)。

1.5 データソース

データソース

「特定された集団」のうち、注目している健康事象が起こる可能性がある人々である、リスク曝露人口(population at risk)の特定が必要。その人たちの情報を把握する方法として、既存資料を使う場合と、独自に調査する場合がある。独自に調査する際、観察研究(後述)では母集団を正しく代表する標本抽出が重要。とくに記述疫学(後述)では標本抽出が不当だと無意味。介入研究(後述)では、普通は標本抽出する必要がない。

リスク曝露人口 疾病の程度を示すための分母。例えば、子宮ガンに注目しているときは、男性は含まない。

既存資料の質 地域相関研究やメタアナリシスでは、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か吟味する必要がある

⁵<https://www.mhlw.go.jp/content/10701000/000945063.pdf>

⁶https://www.mhlw.go.jp/stf/newpage_37876.html

⁷https://www.mhlw.go.jp/stf/shingi/shingi-hosho_126694.html

⁸<https://www.mhlw.go.jp/content/10701000/001215093.pdf>

⁹<https://www.mhlw.go.jp/content/10701000/001149286.pdf>

¹⁰令和6年版は、https://www.mhlw.go.jp/toukei/manual/dl/manual_r06.pdf

Chapter 2

疫学研究のデザイン

研究デザイン

研究デザインには観察研究（記述研究と分析研究）と介入研究があり，目的によって，最適なデザインの研究をすべきである。

- 疫学研究は大きく分けると**観察研究**と**介入研究**に分かれる
- 観察研究では，研究者自身が対象集団に対して意図的に介入し，疾病に関する状態を能動的に変えることはない
- 介入研究では，研究者自身が集団に対して意図的に介入し，能動的に割付けを行って，介入の結果によって疾病改善効果が見られるかどうかを検討する
- 介入研究は，観察研究によってかなりの確からしさと因果関係がありそうなことがわかってからでないと倫理的に実施できないので，アプローチの違いというよりも，段階の違いと考えるべき

2.1 観察的疫学研究のいろいろ

- 記述疫学
- 分析疫学
 - － 自然実験
 - － 生態学的研究（地域相関研究）
 - － 横断的研究
 - － 症例対照研究
 - － コホート研究

2.2 記述疫学

- Descriptive epidemiology の訳語
- Last JM [Ed.] *A Dictionary of Epidemiology 4th Ed.* では, descriptive study (記述研究) と書かれている
- 変数の分布を記述することのみに関心があり, そのためにのみデザインされた研究
- 分母を押さえない限り, 罹患数や死亡数しかわからない
- その研究デザインには因果関係あるいは他の仮説検証を含まない
- 得られたデータは状況把握と仮説構築 (問題発見) に寄与
- 疫学研究の第一段階
- ボンベイでのペスト流行から, 日別の死亡数をカウントして流行曲線を求めた研究など

2.3 分析疫学

2.3.1 自然実験

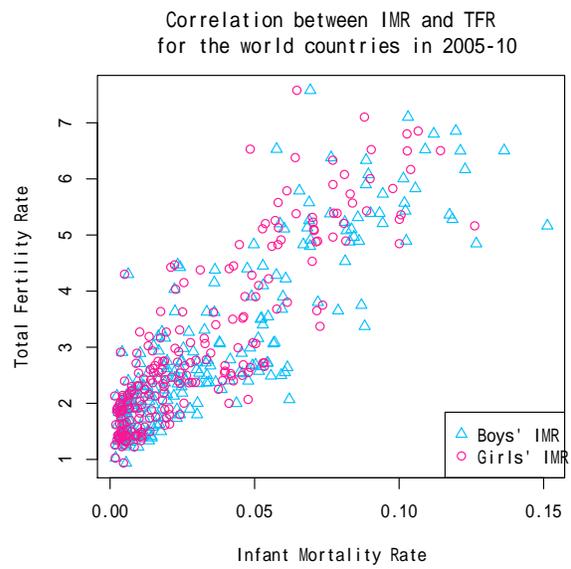
- Natural Experiment の訳語 (下記は Last JM ed., *A Dictionary of Epidemiology 4th ed.* の説明より)
- ある集団に危険因子候補への曝露の水準が異なるいくつかの部分がある, という状況が自然に起こっていて, その状況が, 研究対象となる人々をランダムにグループに割付けた実験的な状況と似ていること (ただし, ある特定のグループにある個人がいるかどうかはランダムではない)
- John Snow がロンドンのコレラ患者数を異なる水道供給会社間で比べたのが有名な例。Snow 自身が “Natural Experiment” と呼んだ
- 他の例としては, 医師の間で喫煙に関連した原因による死亡率が, 同じくらいの年齢の他の専門職の人と比べて低く, それが職業上, 若いうちから禁煙したことに付随して起こった状況など

2.3.2 生態学的研究 (地域相関研究)

- Ecological Study の訳語
- 集団を単位として, 異なる地域に共通する傾向があるかの検討または一つの地域での経時的傾向を調べる (生態学の中ではアレンの法則やベルグマンの法則を想起されたい)
- 交絡因子 (攪乱要因) の影響を受けやすい欠点がある (後述する ecological fallacy がありうる)

- 汚染物質の分布, 汚染物質の食物連鎖, リスク評価などに用いられる

- 変数間の関連性があるかもしれない, という問題発見あるいは仮説構築に寄与



生態学的誤謬の例

ecological fallacy は、通常、生態学的誤謬と訳される。交絡が生じている場合に、集団を単位とすると、個人レベルでの真の関係とは違う関係が見え、間違っただ推論をしてしまうことを指す。例えば下表の地域相関研究データからでは、実際に起こっていることが可能性 1 なのか可能性 2 なのか判別不能

共変量	A 群		計	B 群		計
	X=1	X=0		X=1	X=0	
地域相関研究データ (A 群と B 群で Y=1 となるリスクは同じ)						
Y=1	?	?	560	?	?	560
Y=0	60	40	100	40	60	100
オッズ			5.6			5.6
可能性 1 (X=1 で X=0 に比べ Y=1 となるリスクは 2 倍, X=1 でも X=0 でも A 群で B 群に比べ Y=1 となるリスクは 7/8 倍)						
Y=1	420	140	560	320	240	560
Y=0	60	40	100	40	60	100
オッズ	7.0	3.5	5.6	8.0	4.0	5.6
可能性 2 (X=1 で X=0 に比べ Y=1 となるリスクは 1/2, X=1 でも X=0 でも A 群で B 群に比べ Y=1 となるリスクは 8/7 倍)						
Y=1	240	320	560	140	420	560
Y=0	60	40	100	40	60	100
オッズ	4.0	8.0	5.6	3.5	7.0	5.6

出典: Greenland S (2001) *Int. J. Epidemiol.* 30: 1343-1350.

複数の階層において、どの階層でも個人レベルの関係をみれば 2 つの変数の間に負の相関があるにもかかわらず、階層によって変数の分布範囲に差があるために、全階層をプールした全体としては正の相関があるように見えてしまう状況を考えて、生態学的誤謬が生じる危険性は理解しやすい。

2.3.3 横断的研究

- cross sectional study の訳語。断面研究ともいう
- 対義語は縦断的研究 (longitudinal study)
- 本来の意味は、時間軸と空間軸を考えたとき、1 つの時間で広い空間の断面を切って観察するのが横断的研究。1 つの空間を固定して時間軸に沿って長期間観察するのが縦断的研究。
- 一時点で調査対象が疾病をもっているかどうか、疾病の原因かもしれない要因をもっているかどうかを調べる。得られる情報は有病割合またはオッズ。
- 有病割合は平均有病期間の影響を受ける。疾病負荷の指標といえる。
- 効果指標としてはオッズ比が得られる。

2.3.4 症例対照研究

- 患者対照研究ともいう。case control study の訳語

- ある時点での患者群 (case) に対し、その時点でその疾病をもっていない対照群 (control) を選択し、リスク因子への過去の曝露状況を、患者群と対照群の間で比較するデザイン
- 多くの場合、後ろ向き研究 (retrospective study)
- 対照群の選択が重要。病院対照／健常者対照／一般母集団対照など一長一短。目的次第。
- 因果関係の出口 (病気) を先に押さえて、(患者群と対照群で何の曝露が異なっていたかを明らかにすることで) 何が入口 (要因曝露…複数かもしれない) だったのかを探る研究ともいえる
- 比較的安価で短期間にできる。効率が良い。
- 患者群、対照群それぞれの曝露オッズを求める。
- 効果指標はオッズ比。稀な疾患で、対照をうまく選べば罹患率比やリスク比の近似になる。

2.3.5 コホート研究

- コホート (cohort) とは、ある時点から何らかの共通特性をもった集団として追跡する対象。人口学では同時出生集団をさし、例えば「1980 年生まれ女子コウホート」のように使うが、疫学では共通特性は同時出生に限らない。
- あるリスク因子に曝露した集団を、その後、曝露コホートとして追跡調査 (followup study) し、疾病の発生率を観察。そのリスク因子に曝露していない点だけが異なる非曝露群との比較が理想
- 因果関係の入口 (リスク因子曝露) がわかっていて、(曝露群と非曝露群でどういう病気の発生率に差が出るかを観察することによって) どんな出口 (病気) につながるかを探る研究。
- 疾病量としては、罹患率や累積罹患率 (リスク) が得られる。
- 効果指標は相対危険 (=リスク比や罹患率比) や超過危険 (=リスク差や罹患率差)
- 研究に時間と費用がかかる

2.3.6 より工夫した観察研究

ケースコホート研究 対照が症例と同じコホートから選択されるが、その選択が症例の発症前に行われる症例対照研究。対照群には後に発症する人も含まれうる。ケースコホート研究のオッズ比は、稀な疾患でなくても累積罹患率の推定値となる。

コホート内症例対照研究 (nested case-control study) 追跡中のコホートから発生した患者を症例群とする。同じコホート内の非患者の中から適切な対照群を選択（選択が症例の発症後に行われる）。コホートの過去の情報に遡って症例対照研究を実施。症例も対照も実施中のコホートから得るので、予め定期的に情報が得られていることを利用する。

2.4 介入研究

- 介入研究では、研究者が要因曝露をセッティングすることにより、要因曝露の有無だけが異なる対照群を作り出すことができる。
- 因果関係の入口を決めて出口を評価するという点ではコホート研究に似ているが、曝露の有無をランダムに研究者が割付ける点と、想定される出口が1つである点の特異
- 薬を開発する際の臨床試験 (= clinical trial, 治験ともいう) で盛んに行われる。
- 臨床試験には第1相から第4相まである。
- 中でもRCT (Randomized Controlled Trial; ランダム化統制試験) は、最も科学的に厳密な仮説検定の方法とみなされている。
- 第3相臨床試験では、通常RCTが行われる。

Chapter 3

標本抽出法

標本抽出法

母集団を正しく代表する標本を抽出することが肝要。最も考え方が単純なのはランダムサンプリング (Random sampling, 単純無作為抽出法)。大集団からサンプリングの偏りをなくすためには、多段抽出法、とくに層別無作為抽出法^aを用いる。途上国の調査ではクラスターサンプリング (Cluster sampling, 集落抽出法) がよく使われる (母集団全員の個人名のリストが不要なので)。

^aStratified random sampling。層化無作為抽出法ともいう

疫学調査で得られた結果を適用したい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などの理由で実施されないことが多い。その代わり、集団全体を代表する適当なサイズの標本をうまく選ぶ。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。

3.1 単純無作為抽出法

1. まず母集団の全員をリストし連番を割り振る。
2. 乱数表、さいころ、コンピュータなどを使ってランダムな番号を必要な個数選ぶ。例えば次のようにする。
 - (a) 全員に (0,1) の一様乱数を与える
 - (b) 小さい順に並べ替える
 - (c) 小さい方から必要なところまで対象とする

母集団のリストさえできていれば、ソフトウェアを使うと簡単。例えば、1番から N 番までの番号が振られた人のリストがあって、そこからランダムに p 人分

の番号を抽出したいとき、国際的に広く使われているフリーソフトウェアである R¹を使えば、下記 2 行のどちらでも OK。

```
print(sapply(1:p, function(x, y) which(x==y), rank(I)))
print(sample(1:N, p, replace=FALSE))
```

3.2 層別抽出法（層化抽出法）

まず一段目で年齢、性、職業別など、既知の階層に分け、二段目として各階層ごとに標本抽出を行う方法である（大集団を対象にするときは一段目が居住地域、二段目が年齢と性別、三段目で階層ごとの標本抽出をするなど、三段以上になる場合もある）。階層ごとの抽出が単純無作為抽出であるとき、層別無作為抽出法または層化無作為抽出法と呼ぶ。層によって調査指標が異なることが既知の場合は単純無作為抽出より代表性がいいこと、層ごとの集計ができることが利点である。

反面、サンプリング以前に、階層の情報がわかっていなければならない（が、予備的にその集団について階層を調べたりすると、それ自体が本調査に影響するかもしれない）こと、階層の出現頻度が事前にはわからないこと、時間と金がかかること、全体としてのサンプルサイズが費用や時間などの制約で決まっている場合、階層毎のサンプルサイズが小さくなってしまふことが欠点である。

3.3 集落抽出法

Cluster sampling という英語のまま呼ぶ方が通りが良い。これも多段抽出法のひとつで、最終段の前までは集落を対象とした抽出を行い、最終段では抽出された集落全体を標本とする。母集団全体のリストを作ることが困難な場合に有効な方法である。

途上国で調査をしたいとき、対象とする人々がいくつの村に分かれて住んでいるかはわかっているけれども、それぞれの村にどういう人が何人住んでいるかはわからないのが普通である。このとき、例えば村が 50 個あって、全体の 10% にあたる標本を抽出したければ、5 個の村を無作為に選び、選ばれた村は全数調査（悉皆調査）することによって、対象集団全体を代表する標本を得ることができる。

すべての村から均等に 10% ずつの無作為抽出をするのに比べると、遙かに手間も時間もかからず、同意も得やすいのが利点である。

3.4 確率比例抽出法と副次抽出法

確率比例抽出法とは、Probability Proportionate Sampling (PPS) の訳である。母集団が不均質なとき、均質と考えられるブロックに分け、各ブロックの人口に比例した確率でいくつかのブロックを選んだ後、各ブロックからは同数のサンプルを抽出する。

¹<http://www.r-project.org/> を参照

逆にブロックサイズによらず等確率でいくつかのブロックを選んだ後、各ブロックからそのサイズに比例したサイズのサンプルを抽出する方法を副次抽出法という。

国全体を対象にするなどの大規模な調査では、これらの抽出法を組み合わせることも使うこともある。

3.5 標本抽出法の選択

目安としては以下。

- 母集団が小さいときは、単純無作為抽出か二段層化無作為抽出
- 母集団が大きく資金が乏しいとかアクセスが悪いときは集落抽出
- 母集団が大きいときは、資金が豊富にあれば三段以上の層化無作為抽出または確率比例抽出などの組合せ

3.6 サンプルサイズ

サンプルサイズ

記述研究の最適サンプルサイズは、母集団のサイズ、許容標本誤差、予想頻度から計算できる。分析研究のサンプルサイズは使う統計的検定手法ごとに、予想される測定値のばらつき、臨床的あるいは科学的に意味のある差や相関のレベル、有意水準、検出力を決めて、それぞれの計算式に従って計算できる。ソフトウェアで計算可能だが、費用や時間などの制約により、最適サンプルサイズの研究ができるとは限らない。

- 標本は大きければいいというものではない。最適なサイズが存在する。
- 目的によって考え方が異なる。疾病の頻度や分布を推定することが目的である記述研究の場合は、母集団のサイズ、許容できる標本誤差のレベル（欲しい信頼区間の幅ともいえる）、予想される有病割合（あるいは罹患率、死亡率等）がわかれば計算できる。仮説を検証する（差があるか、関連があるかなど）分析的研究の場合は（観察研究と介入研究のいずれも）、測定値について期待されるばらつき（標準偏差など）、臨床的あるいは科学的に意味がある差や相関のレベル、有意水準、検出力を決めれば計算できるが、統計手法ごとに計算式が異なる。
- どうやって最適サンプルサイズを計算するのは生物統計学のテーマ（後述）だが、通常は統計ソフト（既に紹介した R など）や専用ソフト（Vanderbilt 大学で開発された Power and Sample Size²やドイツで開発された G*Power³など）を用いて計算するのが便利である。

²<http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>

³<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>

- しかし、費用や時間、対象適格性を満たす人が少ないなどの理由で、必ずしも最適サンプルサイズを達成できるとは限らない。その場合は、検出力が足りない可能性を考え、検出力分析を行うべきである。

Chapter 4

疾病量の把握

疾病量の把握

疾病量は、横断的研究では有病割合または疾病オッズ、コホート研究では罹患率や累積罹患率（リスク）として得られる。症例対照研究では疾病量そのものは原理的に把握できない^a。

^a敢えて言えば、研究対象施設における罹患数は疾病量といえるが。

4.1 疾病量の指標

疾病量の指標

横断的研究では有病割合と疾病オッズが得られる。症例対照研究では症例群、対照群それぞれにおける曝露オッズが得られる。コホート研究では罹患率や累積罹患率（リスク）が得られる。

以下3つの区別は重要。

有病割合 (prevalence) ある時点で調べた人のうち病気だった人の割合。無次元。

罹患率 (incidence rate) 観察した人・時間の合計で罹患数を割った値。(／時間) という次元をもつ¹

累積罹患率 (cumulative incidence rate) リスクともいう。観察開始時にいた人のうち、観察期間内に罹患した人の割合。無次元。

- 疾病負荷を示すのは有病割合、発生頻度を示すのは罹患率やリスクである。
- 断面研究（横断的研究）で得られるのは有病割合（と疾病オッズ）のみ。罹患率や累積罹患率を調べるにはコホート研究が必要。記述的研究では、必ず得られるのは罹患数や死亡数（の報告数）のみ

¹単位時間当たりの発生速度を意味する。

4.1.1 有病割合 (prevalence) とオッズ (odds)

- 有病割合やオッズは断面研究（横断的研究）で得られる
- 有病割合は、「ある集団の調査対象者全員」のうち「調査時点で疾病ありの人数」の割合（無次元）
- 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。疾病負荷の指標。
- 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：集団健診をしたら、高血圧や高コレステロール血症の有病割合が高かったので、その対策キャンペーンをすることに決定する、など。
- それに対して、一般に、ある事象が起きる確率の起きない確率に対する比をオッズという。「調査時点で疾病ありの人数」の「その時点で疾病無しの人数」に対する比はオッズである。厳密には疾病オッズという。
- 症例対照研究では、疾病オッズとは違うタイプのオッズが得られる。症例群、対照群それぞれの中で、**曝露を経験していた人数の、曝露を経験していない人数に対する比**もオッズであり、厳密には**曝露オッズ**という。ただし、曝露オッズは疾病量の指標ではない。

4.1.2 罹患率 (incidence rate)

- 個々の観察人時の総和で新規発生患者数を割った値が罹患率である。次元は1/時。単位時間が年ならば分母が人年単位になり、罹患率の単位は1/年。
- 本来はコホート研究が必要。全数報告の疾患では、全人口を1年間観察したと想定し、報告数を中央人口で割った値を罹患率とみなせる。
- 地域がん登録のように、全数報告はされていないが死因統計で補正して罹患率を推定できる場合もある。
- “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／罹患率という関係が成立する（重要!!）

4.1.3 死亡率 (mortality rate)

- 期首人口を1年間追跡して観察される死亡数を**観察人年**で割った値（単位は1／年となる）
- 分母分子ともカテゴリ分けしてカテゴリごとに計算した死亡率はカテゴリ別死亡率 (category-specific mortality rate) となる。死因別死亡率 (disease-specific mortality rate) は分子のみカテゴリ別
- 一般に期間は1年間とするので、分母は1年間の半ばの人口を使い、それを年央人口と呼んで（日本の人口統計では10月1日人口を用いる）、年央人口でその年の死亡数を割って近似する
- 意味：疾病がもたらす結果の1つを示す指標
- 年齢によって大きく異なるので、年齢で標準化することが多い

4.1.4 累積罹患率 (cumulative incidence rate=risk)

- リスクともいう。効果指標のリスク比（後述）は累積罹患率の比。
- 期首人口（観察開始時にいた総人数）のうち観察期間中に疾病に罹患した人数の割合。無次元。
- **観察期間が明示されていないと無意味**
- 追跡調査でしか得られない。脱落者は分母から除外する。
- 無作為割付けの介入研究でよく使われる指標。

4.1.5 致命割合 (case fatality ratio または case fatality risk; CFR)

- （古典的には致命率 case fatality rate と呼ばれてきたが、断じて rate ではないし、最近の疫学者は ratio を使う人が多い。厳密には risk の方が良い）
- ある疾病に罹患し**確定診断がついた人のうち**、その疾病で死亡に至った人の割合（通常、%で表す）²
- 意味：疾病の重篤度を示す。

²西浦らは、2009年パンデミックインフルエンザの流行時に、当初この定義に従って0.4%など高めの値が報告されていたけれども、夏頃から受診しない人や確定診断がつかないままに治療を受けた人を考えると0.045%、0.005%など低い値が相次いで報告され、それがCFRと混同されたために当初報告値が過大評価という批判を受けたことから、確定診断がついた症例中の致命割合 (confirmed case fatality ratio; cCFR) と症状から罹患したと考えられる人の中での死に至った割合 (symptomatic case fatality ratio; sCFR) を区別することを提唱した。2020年のCOVID-19パンデミックに際しては、感染の半分が無症状の人から起こっていることや確定診断にはRT-PCRを要することから、推定された総患者数を分母とするInfection Fatality Risk (IFR)を感染致命リスクと呼び、CFRよりもその方がユニバーサルに使える指標であることが提唱された。

- 例：狂犬病は100%，エボラウイルス病は50～90%，高病原性鳥インフルエンザは60%程度，1918年にパンデミックを起こしたスペイン風邪インフルエンザが2.5%，季節性インフルエンザは0.05～0.1%
- ただし慢性疾患では有病期間が長いので，あまり使われない。
- 致命割合＝死亡率／罹患率
- より厳密には，5年生存率等の方がよい

4.1.6 死因別死亡割合 (proportional mortality ratio; PMR)

- ある特定の死因による死亡が全死亡に占める割合。
- 死因別死亡数の増減はその疾患の増減だけでなく，他の疾患の増減とも連動する（他の疾患による死亡が競合リスク (competing risk) となるため）。
- 例：脳血管疾患による死亡が減ると，心疾患やがんによる死亡が増える
- 割合で見ると，影響はより大きく見える

4.1.7 PMI (proportional mortality indicator) = 50歳以上死亡割合

- 全死亡数に対する50歳以上死亡数の占める割合（%表示）
- 計算に必要なのは年齢2区分の死亡数のみなので，小集団でも信頼性が高い指標
- ただし無文字社会などでは50歳という年齢に意味がない場合もある
- （例）パプアニューギニアのギデラ語を話す人たちは，40年前は，ほとんどの人が自分の年齢を知らなかった

4.2 年齢による標準化

年齢による標準化

疾病罹患や死亡は年齢による影響を受けるので，年齢構造が異なる集団間で比較するためには（例えば年次推移を見たり国際比較するためには），年齢による標準化を行う必要がある。標準化の方法には直接法と間接法がある。直接法の考え方は簡単だが，年齢別の率が必要なのでデータを得にくい途上国では使いにくい欠点がある。

死亡率では直接法年齢調整死亡率と間接法年齢調整死亡率。佐藤，松山 (2011) 交絡という不思議な現象と交絡を取りのぞく解析—標準化と周辺構造モデル—，計量生物学，32Special Issue: S35-S49. も参考になる。

4.2.1 直接法の考え方

- 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
- 基準集団の年齢構成を重みとする，対象集団の年齢別死亡率の重み付き平均ともいえる
- 年齢構成を揃えて考えるために使う基準集団の年齢別人口を i 歳について P_i ，対象集団の年齢別死亡率を m_i と書くと下式の通り。

$$\text{直接法年齢調整死亡率} = \frac{\sum^i m_i \cdot P_i}{\sum^i P_i}$$

- 対象集団の年齢別死亡率の情報が必要

4.2.2 間接法の考え方

- 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た値を，標準化死亡比 (SMR) と呼ぶ。対象集団の死亡総数を d ，対象集団の年齢 i 歳の人口を p_i ，基準集団の年齢 i 歳の死亡率を M_i と書くことにすると，下式で得られる。つまり，対象集団の粗死亡率を，対象集団の年齢別人口を重みとする基準集団の年齢別死亡率の重み付き平均で割った値になる。

$$\text{SMR} = \frac{d}{\sum^i M_i \cdot p_i} = \frac{d / \sum^i p_i}{\sum^i M_i \cdot p_i / \sum^i p_i}$$

- SMR を基準集団の粗死亡率に掛けた値が間接法年齢調整死亡率となる。上記記号を使うと下式の通り。

$$\text{間接法年齢調整死亡率} = \text{SMR} \times \frac{\sum^i M_i \cdot P_i}{\sum^i P_i}$$

- 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だけで計算可能

Chapter 5

危険因子とその影響（効果）の指標

前章の疾病量は、要因ごとに把握し、疾病罹患の危険因子としての要因曝露との関連を調べることによって、要因曝露が疾病罹患に及ぼす影響（効果）を把握することにつながる。

研究前に、測定すべき要因が研究目的にかなうか？ 交絡要因や交互作用をみる要因はないか？ 要因の対象集団内でのばらつきは十分か？ 信頼できる測定方法はあるか？ を吟味し、既存資料、個人の提供情報、医学的検査・測定、環境測定などから適切な情報源を選んで情報を得る。

効果の指標

危険因子曝露の効果を示す指標としては、横断的研究と症例対照研究ではオッズ比、コホート研究や介入研究では罹患率比やリスク比、死亡率比を用いることができる（ただし介入研究の場合はコックス回帰でハザード比を計算する方が普通）。

5.1 危険因子とは？

- 「あるリスクをもたらす要因」のこと
- もう少し限定的な定義：「疾病の発生あるいは他の特定結果の起こる確率を増加させる属性または曝露を危険因子と呼ぶ」
- この確率の増加を、その危険因子の影響（または効果）と呼ぶ

5.2 影響（効果）の指標

危険因子がある群と危険因子がない群の間で疾病量の差や比（または変化率）をとるとするのが基本的な考え方。

超過危険 (excess risk) 寄与危険 (attributable risk) ともいう。曝露群のリスクや罹患率から非曝露群のリスクや罹患率を引いた差。コホート研究と介入研究で得られる。リスクの差はリスク差、罹患率の差は罹患率差、死亡率の差は死亡率差。公衆衛生的なインパクトを見るのに適している。統計学的に0と有意に異なれば曝露は罹患に効果ありといえる。

相対危険 (relative risk) 危険因子への曝露群の非曝露群に対するリスクの比（リスク比）、罹患率の比（罹患率比）、死亡率の比（死亡率比）の総称。コホート研究と介入研究で得られる。因果関係を検討するのに適している。公衆衛生的なインパクト評価としては過大な可能性。統計学的に1と有意に異なれば曝露は罹患に効果ありといえる。

オッズ比 (odds ratio) 断面研究では要因あり群の疾病オッズの要因なし群の疾病オッズに対する比（疾病オッズ比）、症例対照研究では患者群の曝露オッズの対照群の曝露オッズに対する比（曝露オッズ比）をいう。下記の通り両者は数学的に一致する。稀な疾患では相対危険の近似値となる。症例対照研究では、対照群を密度依存サンプリングするなど工夫すれば、オッズ比が理論的に率比に一致する。

クロス集計表を作るとわかりやすい。

	病気あり	病気なし
曝露あり	a 人	b 人
曝露なし	c 人	d 人

このとき、断面研究における¹疾病オッズ比 (disease odds ratio) は、

$$\frac{(a/b)}{(c/d)} = \frac{ad}{bc}$$

症例対照研究における曝露オッズ比 (exposure odds ratio) は、

$$\frac{(a/c)}{(b/d)} = \frac{ad}{bc}$$

となって一致することが自明である。

5.2.1 オッズ比の計算例

『シンプル衛生公衆衛生学 2023』でも紹介されている Doll and Hill (1950) の BMJ 論文²のデータを例にあげる。

ロンドンの 20 か所の病院で、1948 年 4 月から 1949 年 10 月のがん診断例のうち 75 歳未満の肺がん患者 709 名（男性 649 名、女性 60 名）を症例、性別と年齢をマッチングしたがん以外の患者 709 名を対照として喫煙との関係を分析したもの³

¹コホート研究でも考えることは可能だが、コホート研究ではリスク比や率比を計算できるので、敢えてオッズ比を計算する理由はないだろう。

²<https://doi.org/10.1136/bmj.2.4682.739>

³症例対照研究でのマッチングなので、対照が母集団を正しく代表しておらず効果指標が過小評価になる潜在的バイアスがある点に注意。

	肺がん	肺がん以外
喫煙者	男性 647, 女性 41	男性 622, 女性 28
非喫煙者	男性 2, 女性 19	男性 27, 女性 32

全体でのオッズ比は

$$\frac{(647 + 41)/(2 + 19)}{(622 + 28)/(27 + 32)} = 2.97$$

男性のオッズ比は

$$\frac{647/2}{622/27} = 14.04$$

女性のオッズ比は

$$\frac{41/19}{28/32} = 2.47$$

男女を層別しないと男性のリスクをかなり過小評価することになる。

5.3 寄与割合・人口寄与割合

寄与割合 (Attributable Proportion) 曝露群の罹患率のうちその曝露が原因となっている割合。つまり罹患率差を曝露群の罹患率で割った値。罹患率比から1を引いて罹患率比で割った値とも等しい。

人口寄与割合 (Attributable Population=Attributable Fraction) 母集団の罹患率のうちその曝露が原因となっているものを取り除くとどれくらいの割合、罹患率を下げられるか？ という意味をもつ。『シンプル衛生公衆衛生学 2023』では、集団（人口）寄与危険割合 (population attributable risk fraction = PAF) と書かれている。計算式としては、罹患率差を母集団の罹患率で割った値になる。

5.4 相対危険と超過危険の関係

相対危険が1で超過危険が0という状態を除けば、相対危険が経時的に不変、即ち曝露群のリスクと非曝露群のリスクの比が年齢によらず一定 (> 1) であるとしたら、加齢に伴ってリスクそのものが大きくなることによって曝露群と非曝露群のリスクの差が拡大していくので、超過危険が加齢とともに増加する。逆に、超過危険が経時的に不変、即ち曝露群と非曝露群のリスクの差が年齢によらず一定であるとしたら、やはり加齢に伴ってリスクそのものが大きくなることによって、相対危険は減少していく。

5.5 生涯リスク

- 生涯リスクとは、ある要因 X による超過危険の、生涯に渡る積算値である。

- PYLL (Potential Years of Life Lost)：リスクの増加に伴って失われる余命の指標。所与の生存目標年齢に達する前に死亡した場合に死亡年齢と生存目標年齢の差を求め、それを合計した値。
- ゼロ歳損失余命 (Loss of Life Expectancy at Age Zero)：生存目標年齢を平均寿命にした場合の PYLL の一つ。

Chapter 6

因果関係

6.1 因果関係とは？

- 要因（リスク因子）→影響（病気）の関係¹
- 完璧な証明は困難
- 統計学的な相関関係は何らかの関連を示唆する……が、
 - 因果の向きは不明。
 - 無関係なのに偶然相関がでてしまったかもしれない（第一種の過誤）
 - 見かけの関連かもしれない
 - バイアスや交絡（後述）のせいかもしれない

もっとも強い因果関係は、要因が影響を起こすメカニズムが生物学的に明らかであり、それが常に成り立つ場合にいえる（生物学的因果関係）。疫学の究極の目的は、それを明らかにすることともいえる。

6.2 第一種の過誤・第二種の過誤

統計学的な相関関係が偶然とは考えられないほど大きいかどうかは、ネイマン＝ピアソン流の仮説検定で行うのが普通である。通常、母集団において「差が無い」あるいは「関連が無い」という帰無仮説を立て、一方で「AがBより大きいまたは小さい」とか「Aの死亡率はBの死亡率より高い」という対立仮説を立てる。

両方を立てて統計的検定を実行した結果、 p 値 < 有意水準ならば、帰無仮説を棄却し対立仮説を採択するという意思決定を行う。一方、 p 値 \geq 有意水準のときは、帰無仮説を棄却しないで保留する（積極的採択ではないことに注意）。

このとき、本当は帰無仮説が正しい（母集団では「差が無い」「関連が無い」）のに、誤って対立仮説を採択してしまう確率は、有意水準に等しい。この意味で、有意水準は**第一種の過誤（ α エラー：アワテモノのエラーと覚えると良い）**

¹その要因が原因となって、結果としての影響が表れる、という意味。

逆に、本当は「差がある」「関連がある」のに、サンプルサイズが小さいなどの理由で帰無仮説を棄却できない確率を、**第二種の過誤（ β エラー：ボンヤリモノのエラーと覚えると良い）**と呼ぶ。（1 - 検出力）に等しい

6.3 生物学的因果関係を導く指針（不完全）

生物学的因果関係の存在を示す指針としては以下のものが知られている。しかし時間性を除きすべて例外がある。

- 科学的常識
- Henle-Koch の 4 原則【(1) ~ (3) を 3 原則という】
 - (1) その病原体が当該感染症患者から分離される
 - (2) その病原体は他の疾病患者には見出されない
 - (3) 患者から分離培養された病原体が実験動物に同一疾患を発生させる
 - (4) 当該罹患動物から再び同一の病原体が分離される
- 動物実験・実験室の事実
- 病理学的事実
- 観察疫学的事実（ただし、**時間性以外はすべて例外があることに注意**）²
 - (1) **Strength** 関連の強さ。喫煙していない医師の死亡率が 0.07/1000/年、一日 1-14 本喫煙する医師の死亡率が 0.57/1000/年、一日 25 本以上喫煙する医師の死亡率が 2.27/1000/年というデータから、死亡率が 1-14 本の喫煙で 8 倍になり、25 本以上の喫煙で 32 倍になるといえる。喫煙と死亡の間には強い関連があると考えられる。
 - (2) **Consistency** 一貫性。その関連は異なる人、異なる場所、異なる状況、異なる時点でも一貫して見られるか？
 - (3) **Specificity** 特異性。特定の作業に従事する労働者に特定の部位の疾病が起こりやすいという関連性がある、その部位の疾病が他の労働者には稀であるならば、その作業への従事はその疾病の原因であるという因果関係を考える根拠となりうる。
 - (4) **Temporality** 時間性。原因は必ず時間的に結果より前に起こっているはずで、これは例外がない。
 - (5) **Biological gradient** 用量反応関係が成立しているか。それが真に病気の原因であれば、曝露の量が多かったり頻繁なほど、その病気にかかる確率は高まるはず、と考えられる。

²<https://www.mhlw.go.jp/shingi/2005/03/d1/s0331-9b2.pdf> や <http://www.edwardtufte.com/tufte/hill> を参照。Sir Austin Bradford Hill が “The Environment and Disease: Association or Causation?,” *Proceedings of the Royal Society of Medicine*, 58: 295-300, 1965. で提示した、通称 Hill の基準。ただし、Rothman は基準ではなくチェックリストと呼ぶべきと警告している。

- (6) **Plausibility** 蓋然性。もっともらしさ。
- (7) **Coherence** 整合性。他の知見と矛盾しないか。
- (8) **Experiment** 実験あるいはそれに準ずる方法で証明できるか。
- (9) **Analogy** 似た現象から類推できるか。

- 介入研究・実験疫学・臨床試験による
- メタアナリシスで複数の研究をまとめて解析→共通して同じ関係検出

6.4 個人レベルでの因果関係は立証可能か？

- 例えば、中澤は、ソロモン諸島の主なマラリア媒介蚊が踝から下を吸血するので、ソロモン諸島で調査するときは、ずっと靴下をはいている。
- いまのところマラリアに罹ったことはないが、それが本当に靴下をはいているから吸血予防ができていて罹らないのか、それとも、仮に靴下をはいていなくてもマラリアには罹らなかったのかは、**靴下をはかなかつた中澤が存在しない以上**、わからない。
- つまり、個人レベルでの因果関係は立証不可能!!
- こういう考え方を**反事実 (counterfactual)** モデルという。
- つまり、目の前の患者さんの病気について、因果関係を確定することはできない。できるのは、正しい可能性が高い因果推論 (causal inference) をすることだけ。

6.5 集団レベルでの因果推論のロジック

- 喫煙していて肺がんにかかった A さんが、もし喫煙しなかったら、という反事実は観察不能
- *しかし* 喫煙という曝露 (exposure) 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 は設定可能
- ということは、集団 A と集団 B を追跡し、集団間で肺がんの発生率を比較することも可能
- ⇒集団レベルの因果関係がいれば、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、曝露群と非曝露群 (対照群)
- 「喫煙という曝露条件以外の条件がほとんど同じ」になるように非曝露群を選ぶ操作をマッチング³と呼ぶ。このようなコホート研究では、マッ

³細かくいうと、個人単位のマッチングと集団レベルのマッチングがあり、個人単位の場合も曝露群 1 人に対して非曝露群が 1 人だったり複数人だったりするなど、いろいろなやり方がある。

グは背景因子を揃える（曝露の有無以外の条件がほとんど同じと仮定できる）ことになるので有益。ただし、症例対照研究では、患者群に対して対照群（非患者群）をマッチングすると、母集団を正しく代表しなくなる危険があるので、必ずしも勧められない。

6.6 いろいろな因果関係

6.6.1 一要因⇒一症状

(例) 交通事故⇒外傷による死亡。

6.6.2 一要因⇒多症状

(例) 鉛の吸収過剰：腎臓，肝臓，中枢神経系や骨に蓄積し，食欲不振，筋肉の痛み，腹痛，不妊，脳疾患（鉛エンセファロパシー），慢性腎炎などを引き起こす

6.6.3 多要因⇒一疾病

(例) 肝細胞がんが発生するには，B型肝炎やC型肝炎への罹患，多量飲酒や喫煙など，多くの要因が積み重なることが寄与する。

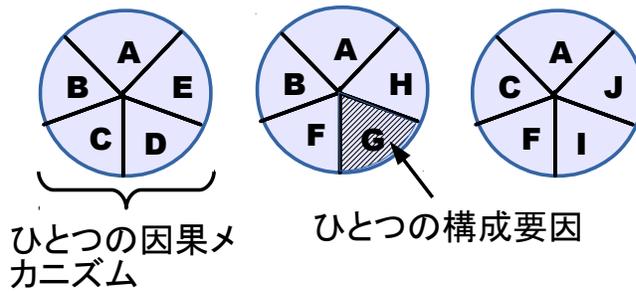
cf. <http://www4.ocn.ne.jp/~etrtr/18yma.htm>

6.6.4 多要因⇒多疾病

(例) 糖尿病や冠動脈疾患など，多くの慢性疾患は，高血圧，メタボリックシンドローム，不摂生，ストレスなど多くの要因が複雑に絡み合っ発症に至っている。

6.7 因果関係の整理～因果パイモデル

- 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）



6.7.1 因果パイモデルの効用

ある疾病にかかわるすべての因果パイがわかれば、以下の利点がある。

- すべての因果パイを相互に比較する
- 各要因の**相対的重要性**を評価できる⁴

1つの因果パイについてみると、

- 最初の要因が作用してから、**最後の要因が作用して発症に至るまでの時間**（誘導期間=induction period）を評価できる
- 感染症に比べ慢性疾患では一般に誘導期間が長い

⁴一つの考え方としては、多くの因果パイに含まれる要因ほど重要と考えられる。

Chapter 7

因果関係を歪めるもの～測り間違い

誤差

交絡と誤差が因果関係を歪める。誤差には**偶然誤差**と**系統誤差**がある。偶然誤差には測定機器の精度の低さ、サンプルサイズの小ささからくる標本誤差等が含まれる。系統誤差には**選択バイアス**と**情報バイアス**が含まれる。偶然誤差が大きいと、真値の推定値としての測定値の信頼区間の幅が広くなり、系統誤差が大きいと、測定値の真値とのズレが大きくなる。

7.1 何が因果関係を歪めるのか？

因果関係はさまざまな要因によって歪められ、見えにくくなっていることが多い。多くの場合、注目している要因以外の要因の作用によって関連が歪められている（交絡が存在する）だけでなく、測定や評価が正しくないこと、つまり誤差の影響を受ける。

誤差にはランダムな誤差（偶然誤差）と系統誤差（バイアス）がある。この2つを区別し、かつ、どういうメカニズムで誤差が起こるのかを整理しておく必要がある。

ランダムな誤差が大きい測定とは、精度の低い測定ともいえるが、統計学的には、測定により得られた推定値の信頼区間の幅が広い測定である。検査機器や技術の改善によって減らせるし、サンプルサイズを大きくすることでも減らせる。

系統誤差（バイアス）は、個々の測定値が真値から同じ向きに偏ることである。ゼロ点の合っていない機械による測定値は、いくら精度が高くてもバイアスが大きくなるし、バイアスがあったら、いくらサンプルサイズを大きくしても真値からの推定値のズレは小さくならない。バイアスには、大別すると、**選択バイアス**（selection bias; 観察対象が母集団を正しく代表しないこと）と、**情報バイアス**（information bias; 観察対象から得る情報が偏ってしまうこと）がある。

7.2 選択バイアス

母集団を正しく代表する観察対象を選択できていないことをいう。

7.2.1 Neyman's バイアス

(例1) 喫煙は肺がん罹患リスクを上げるといえる？

- 数値例：肺がん死亡 100 人と他死因 100 人で過去の危険因子曝露を比べる症例対照研究で肺がん死者中 90 人が喫煙者、他死因の死者中 50 人が喫煙者
- オッズ比： $(90/10) / (50/50)=9$
- 喫煙は肺がん死亡リスクを 9 倍にする！
- しかし、肺がん罹患時の致命割合が、喫煙者では 90%、非喫煙者では 5%だったとしたら、罹患リスクはどちらが高いかわからない（肺がん死者が、肺がん罹患者のうち、喫煙者に大きく偏っているため）

(例2) 喫煙はアルツハイマー病を予防する？

- アルツハイマー患者と非患者の 2 群間で、喫煙状況を比べた症例対照研究 8 つをまとめたメタアナリシスで、要約オッズ比は 0.78 (95%信頼区間は 0.62-0.98)
⇒喫煙者はアルツハイマーになりにくい、と結論 (Graves et al., 1991)
- 本当？
- 加齢がアルツハイマーの真のリスク因子で、喫煙者の方が短命ならば、アルツハイマー患者群に喫煙者は含まれにくい！ だから、この結論が正しいとは限らない。

この 2 つの例は、どちらも Neyman's Bias (または Prevalence-incidence bias) と呼ばれる選択バイアスである (Hill et al., JCE 56: 293-, 2003)。

7.2.2 逸話的情報とサンプリングバイアス

オーケストラ指揮者は長生きか？ (出典：Rothman, 2012)

- Boston Globe の特集記事：「有名なオーケストラ指揮者の多くが長生き」
⇒「オーケストラ指揮は健康に良い」
- 本当？
- 確かに、Otmar Suitner (2010 年 1 月 8 日没, 享年 87 歳), Leonard Bernstein (1990 年 10 月 14 日没, 享年 72 歳), Herbert von Karajan (1989 年 7 月 16 日没, 享年 81 歳), Karl Böhm (1981 年 8 月 14 日没, 享年 86 歳) など、長生きした人が多いように感じる。
- しかし実は、彼らは長生きしたから有名になった<逸話的情報>。早世した人が見過ごされた。有名な指揮者は指揮者を代表しない「選択バイアス」

- ……が、それだけ？
- 過去 100 年の指揮者全員の平均死亡年齢が同時期の一般人のそれより高ければ、指揮は健康に良いといえる？
- 実は、指揮者になれるのは音楽キャリアを積んだ後
- 指揮者というだけで、低めにみても 30 歳以上
- 同じ年齢の一般人と死亡リスクを比べないと無意味。つまり、オーケストラ指揮という要因に曝露していない対照（非曝露）群の方を、年齢を限定して標本抽出しないと、サンプリングバイアスが起こってしまう。

7.2.3 調査におけるサンプリングバイアス

新聞社のアンケート結果によくある問題点

- 平日昼にランダムディジットダイヤリング方式 (RDD) で世論調査して、一般母集団を代表できるのか？
- 1000 人から回答が得られるまで電話したとしながら、全部で何回の電話をかけたか書かれていない（固定電話をもっていて、平日昼に在宅で、質問に答える暇がある人しか対象にならないので、おそらく 2000 回以上の電話をかけているはずだが、それを書く回答者の代表性に疑念を抱かれるので書かない例が多い。
- 男女何人ずつ、という形で回答を集めた場合は、おそらく男女で年齢層が異なる

谷岡一郎『社会調査のウソ』（文春新書）に載っている例

- 阪神淡路大震災から約 1 年後の仮設住宅居住者 1000 人に調査した結果と、さらに半年後に、そのうち 300 人に調査した結果を比べると、「復興から取り残される」> 7 割, 「行政に不満」～ 8 割に「増加」⇒何を意味？
- 半年後の調査対象 300 人が仮設住宅に居住し続けているとしたら、被災者全体を正しく代表していない。まさに行政サポートから取り残された人たちに偏っているので、不満があっても当然。
- これもサンプリングバイアスといえる。避けるためには、1 回目調査した 1000 人を追跡して（そこからランダムサンプルしても良いが）調査しなくてはいけない。

7.2.4 Berkson バイアス

胆嚢炎と糖尿病の関係？

- 糖尿病患者を症例、同じ病院に入院している非糖尿病患者を対照として、胆嚢炎の有無を調べる

- 胆嚢炎に罹った人は罹っていない人より入院しやすいので、健康な人を対照とした場合に比べて、胆嚢炎と糖尿病の関係は薄まる可能性がある
- 対照が一般母集団を代表しない選択バイアスであり、入院率バイアス (Berkson's bias) と呼ぶ

7.2.5 自己選択バイアス

- 骨粗鬆症予防の新しい運動プログラムを開発し、参加したボランティアと参加しなかった住民を 10 年間観察し、ボランティアの方が骨粗鬆症発生率が低かったとする
- この運動プログラムは骨粗鬆症予防に有効だったか？
- 運動プログラム参加者は、そもそも健康に関心が高い人で、このプログラム以外にも、カルシウムを十分にとるとか、健康的なライフスタイルをとっている人に偏っている可能性がある。これを**自己選択バイアス**と呼ぶ

7.2.6 健康労働者効果 (Healthy Worker Effect)

- 原発労働者の心疾患罹患率が一般住民の心疾患罹患率より低かったことから、原発作業は心疾患を予防するという説がある。低用量放射線曝露によって免疫系が活性化され、より健康になったという「放射線ホルミシス」を唱える人がいる。
- しかし、「放射線ホルミシス」など考えなくても、この現象は、重労働ができる人は、一般住民に比べて、元々健康な人に偏っているので、原発作業と心疾患が無関係でも予防効果があったように見えてしまう場合がある「**健康労働者効果**」で説明がつく。
- これもよく知られた選択バイアスの一つ。

7.3 情報バイアス

観察対象から正しく情報をとりだせていないことをいう。

7.3.1 追跡の偏り

肺気腫への喫煙の影響を調べるコホート研究の例

- 喫煙者と非喫煙者を 1000 人ずつ 10 年間追跡。喫煙群では延べ 100 人、非喫煙群では延べ 10 人が肺気腫発生と診断されたとする
- 喫煙は肺気腫発生リスクを 10 倍に高めたといえるか？
- 医師は、喫煙者が来院すると必ず肺気腫を疑って慎重に呼吸器系の検査をし、非喫煙者のときは問診で疑いが高い 10% の人だけ慎重に呼吸器系検査をするかもしれない。その場合は、肺気腫の発生リスクに差が無くても、上記診断率の差はでてしまう。これを「追跡の偏り」という。

7.3.2 リコールバイアス

後ろ向き研究では、過去の事実を思い出して貰って情報を得るが、人間の記憶はしばしば間違ふ。この、記憶を想起して貰う際の誤りを「リコールバイアス」と呼ぶ。間違いかたに曝露の有無によって差があることが大きな問題。

母性の思い出しバイアス (maternal recall bias) の例

- 先天異常をもつ子どもを症例、先天異常が無い子どもを対照とする症例対照研究
- 母親に妊娠初期の風邪薬 A の服用について尋ね、症例の方が高い割合で、A を「服用した」と回答。
- A は先天異常の原因といえる？
- 症例群の母親の方が、出産直後の時点で、過去について真剣に原因を考えて思い出すので、仮に A の真の服用割合が同じでも、対照群の母親より服用経験を思い出す割合は高い

7.4 交絡 (confounding)

交絡とは、注目している因果関係を歪める、それ以外の因子である。

7.4.1 交絡要因の3つの必要条件 (例：肥満の高血圧に影響するという因果関係における年齢の関与)

- 注目している要因 (肥満) ではない
- 注目している要因 (肥満) と結果である健康影響 (高血圧) の両方と因果的に関連 (注：「結果に影響する要因」かつ「比較したいグループの一方に偏っている要因」とも言える)
- 要因曝露の結果ではない (肥満が高齢をもたらすのではない) (注：「原因と結果の間にある中間変数ではない」とも言える)

制御すべき交絡を見つけるには、DAG を使った裏口テストが有用である。日本製薬工業協会シンポジウムの資料¹が参考になる。

7.4.2 交絡要因を見過ごすと変な因果推論をしてしまう

例：スウェーデンの女性はカザフスタンの女性より粗死亡率が高い

- スウェーデンとカザフスタンの女性の 1992 年の死亡率を比べる
- 総死亡数を総人口で割った「粗死亡率」は医療水準が高いと思われるスウェーデンの方がカザフスタンより高い。

¹https://www.jpma.or.jp/information/evaluation/symposium/pdf/20210319/2_2.pdf

- 実は、年齢構造が交絡要因となっている。
 - 年齢階級別死亡率は、どの階級でもカザフスタンが上
 - しかし、粗死亡率はスウェーデンが人口千対 10.5、カザフスタンが人口千対 6.3 と、スウェーデンが高い
 - 原因は、スウェーデンの方が高齢者が多く、スウェーデンでもカザフスタンでも高齢者の死亡率は若い人より遙かに高いこと。

7.4.3 交絡の制御

交絡を制御するには、**デザインによる制御**と**分析段階での制御**がある。例えば、性別や年齢による交絡がありそうな因果関係についての研究を行う場合に、性別と年齢で層別した**層別無作為抽出**をすることや、コホート研究であれば曝露群と非曝露群について性別と年齢の**マッチング**を行うことや、特定の性別・年齢に**限定**して研究を実施することは、**デザインによる制御**である。

分析段階の制御には、下記のように、層別解析、標準化、プール化、多変量解析がある。

層別解析 交絡因子によって層別し、別々に解析すること。年齢による交絡があるなら、年齢階級別に死亡率を比べてみるなど

標準化 交絡因子について共通の重みを設定し、標準化した指標を求めて比較すること。例えば、年齢調整死亡率など。

プール化 交絡因子によって層別した各層に共通の関連性がみられると想定し、関連性を要約すること。マンテルヘンツェルの要約オッズ比など。

多変量解析 統計学的に、説明変数群の中に、交絡が疑われる変数を共変量として投入することによって、交絡の影響を制御することができる。ロジスティック回帰分析などがよく行われる。

Chapter 8

疫学研究の歴史と研究倫理

8.1 古典的疫学研究

- 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。
- Snow J (1813-1858) のコレラの研究：当時コレラ菌は未知だったが、ロンドンではコレラが日常的に流行していた。1854 年の大流行時に、John Snow は、2 つの研究によって、飲み水によってコレラが伝播している可能性が高いことを示した。1 つは、コレラ死亡者の発生地図 (spot map) と発症日別のコレラ死者数の度数分布を作成した記述疫学研究¹により、コレラ流行の原因がブロードストリート地区の 1 つの共同井戸利用にあったことを突き止めたこと。もう 1 つは、水会社間の供給人口当たりのコレラ死亡数の比 (L 社：461/173748 に対して S&V 社：4093/266516 が約 5.8 倍) を計算し、テムズ川下流域から取水している S&V 社から水供給を受けている人がコレラに罹りやすいことを示した自然実験研究。

¹R の HistData パッケージを使えば、library(HistData); example(Snow) によってそれらのグラフを得ることができる。

8.2 代表的な疫学研究

8.2.1 Frammingham Heart Study (フラミンガム研究)

- 米国東部マサチューセッツ州ミドルセックス郡フラミンガム町
- 人口：65598 人（2005 年）
- 1948 年（米国民の死因の 8 割が循環器疾患だった頃）開始
- 疫学の手法を循環器疾患に適用した最初の研究
- オリジナルコホート（1948 年）5209 人，子供世代コホート（1971 年）5124 人。第三世代コホート（2005 年）4095 人。それぞれはクローズド・コホート。町全体としてみればオープン・コホート。
- 検査項目：
 - （オリジナルコホート 1998 年時）身体測定，血圧，尿検査，生活習慣，病歴聴取，MD 検査，生化学検査，心電図，X 線撮影，肺機能検査，身体機能，認知機能，心理社会的状態
 - （子供世代コホート）X 線撮影を除く上記項目＋心エコー，頸動脈エコー，ホルター心電図
- <https://www.framinghamheartstudy.org/>

8.2.2 久山町研究

九州大学のグループが実施した大規模コホート研究²。

- 福岡県糟屋郡久山町（人口 8075 人：2007 年 1 月 1 日）
- 1961 年～，40 歳以上の男女を対象としたコホート研究
- 登録数：第 1 集団 1621 人（1961 年～），第 2 集団 2038 人（1974 年～），第 3 集団 2637 人（1988 年～），第 4 集団 3500 人（2002 年～）
- 定期健診（通常毎年＋一斉 5 年毎）：アンケート調査，食事調査，身体測定（身長，体重，皮下脂肪厚，握力），尿検査（pH，蛋白，糖，潜血，ケトン体，ウロビリノーゲン），血圧測定（坐位，臥位），血液検査（肝機能等 31 項目），貧血検査，糖負荷試験，心電図，眼底検査，問診
- 特徴：全国平均とほぼ同じ年齢・職業分布～平均的な日本人。剖検率が高い（1964 年 100%，その後も 80%），追跡率 99%，一斉健診 78% 受診
- 1961 年当時，日本は脳卒中が死因の 1 位。欧米と違って脳出血が脳梗塞の 12.4 倍。欧米の研究者から死因への疑いが呈され，剖検により正しい死因を確定する目的で開始。後に心血管病，ゲノム疫学へ展開

²http://www.epi-c.jp/e001_1_0001.html

- 主な結果：脳出血による死亡は脳梗塞による死亡の1.1倍（第1集団）。脳卒中の最大の危険因子が高血圧であることを解明，保健指導（全ての健診結果が出た時点で医師が個別指導）＋降圧剤服用で脳卒中死亡を大幅に減らした介入成果，等

8.2.3 INTERSALT 研究

INTERSALT 研究は，個人ベースで多くの国と地域について実施された横断的研究だが，地域集団ごとに集計した値を単位とした地域相関分析によってもさまざまな関連性が示されている。³

- 32か国 52集団（うち日本3集団），20-59歳の男女，1集団 200例ずつ，1985年に横断的に実施
- 登録数 10079例（男性 5045例，女性 5034例）
- 調査項目 血圧（5分間の安静ののち，ランダムゼロ血圧計を用いて座位で2回測定），24時間蓄尿（尿量，ナトリウム，カリウム，マグネシウム，カルシウム），心拍数，運動状況，飲酒，喫煙，服薬状況，年齢，身長，体重，人種，教育，経産歴（女性のみ），閉経状況（女性のみ）。
- 主な結果：地域相関で，BMI \propto 血圧，Na/K比 \propto 血圧，塩分摂取量 \propto 血圧，ヤノミなど低塩分摂取集団では加齢に伴う血圧上昇なし，等

8.2.4 代表的な症例対照研究

- SMON についての椿忠雄の研究⇒キノホルム服用が原因
- 四肢の全部あるいは一部が短い奇形をもつ新生児の多数発生⇒サリドマイド服用が原因

8.3 人を対象とする生命科学・医学系研究に関する倫理指針

情報源としては、https://www.lifescience.mext.go.jp/bioethics/seimeikagaku_igaku.html を参照すると良い。

- 2002年に文部科学省と厚生労働省が合同で、「疫学研究倫理指針」を発表し，2007年全面改訂され，2008年12月にさらに一部改訂された。その後，2014年12月22日に臨床研究に関する倫理指針と統合されて「人を対象とする医学系研究に関する倫理指針」となり，2017年2月28日に一部改正され⁴，さらに2021年からヒトゲノム・遺伝子研究倫理指針と一体化して「人を対象とする生命科学・医学系研究に関する倫理指針」となった。医学研究倫

³http://www.epi-c.jp/e114_1_0001.html

⁴<http://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagaku/0000153339.pdf>

理についてのヘルシンキ宣言や患者の権利についてのリスボン憲章に依拠している。最近の改正は個人情報保護法⁵との関連が大きい⁶。

- 疫学研究は人間を対象とするので、倫理面での配慮が不可欠
- とくに介入研究では曝露条件をセッティングするので、十分に統制された実験をする必要がある
- 観察的研究や記述疫学研究であっても、プライバシーへの配慮が必要
- 文書によるインフォームドコンセントが多くの場合に必要（侵襲を伴わない場合は不要とされるが、得ておくほうが無難）
- 倫理審査委員会による審査を通らないと研究できない（成果を学術雑誌に投稿しても事務的に受け付けを拒否されてしまう）

ただし、システマティックレビューやメタアナリシス⁷、二次資料の解析（例えば、Tsuda T et al. *Epidemiology*, 27(3): 316-322, 2016.⁸）のように、倫理審査を要さない疫学研究もある。

⁵<https://www.ppc.go.jp/personalinfo/>

⁶厚生労働省の説明資料:<http://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000170955.pdf>

⁷詳細は PRISMA 声明 (<http://www.prisma-statement.org/>) 参照

⁸<https://doi.org/10.1097%2FEBE.0000000000000385>

Chapter 9

実験計画と生物統計学

9.1 実験計画における Fisher の 3 原則

反復 1つの処理に対して少なくとも2回以上の繰り返しが必要

無作為化 実験の順序や空間的にどの場所にどの実験群を割り当てるのかを無作為に決める必要がある。

局所管理 実験が大規模で、実験全体を無作為化するのが妥当でないとき、実験をある程度細分化してブロックを構成し、ブロック内で処理条件を無作為化し、ブロック内のバックグラウンドが均一になるよう管理すると、系統誤差の一部がブロック間変動として除去できる。

9.2 実験計画法の発想

ザルツブルグの本（“The lady tasting tea”）に紹介されているエピソード。

- Cambridge の午後、ミルクティーを作るときにミルクを先にカップに入れたのか、紅茶を先にカップに入れたのかを、飲んでみれば見分けられる「ミルクティー判別能力」があるという女性
- 本当かどうか名だたる学者が大論争
- 実験してみれば？ といったのが R.A. Fisher
- どちらを先にして作ったのかを知らせずに、この女性にミルクティーを飲んでもらって当てさせてみれば、本当にミルクティー判別能力があるのかわかるという理屈
- どういう順番で何回試してみれば、得られた結果からその女性にミルクティー判別能力があるのかが判定できるような条件を考える必要がある→実験計画が必要!!

本当はミルクと紅茶の順番で味は変わるのか？

- 「動物農場」や「1984年」などのディストピア小説で知られる George Orwell は紅茶好きとしても有名で、“11 rules for perfect tea making” という文章を発表している。その 10 番目で、「ミルクを紅茶に入れるのだ。逆ではいけない」と書いている。
- 英国人は紅茶が好きで、The Royal Society of Chemistry 主催の会でのスピーチ (2003) で Dr. Andrew Stapley は、「ミルクを先にカップに入れておいて、後から紅茶を注ぐのだ。低温殺菌の冷たいミルクは、後から注ぐとタンパクが変性しやすくまずくなる。先に入れておけば瞬間的な高温に曝されず美味しい」と述べた。(以上 2 つの出典は、BBC のサイト及び RSC のリリース^{a)})
- 日本人でも実験した人がいて (紅茶 130cc + タカナシの低温殺菌牛乳 30cc)、主観的な意見だが、やはりミルクが先の方が美味しい、とブログに書いている (ただし日本で普通に売られている高温殺菌の牛乳では違いが分からないとのこと)^{b)}

^{a)}<http://news.bbc.co.uk/2/hi/uk/3016342.stm>

^{b)}http://blog.livedoor.jp/teatime312/archives/cat_123365.html

9.2.1 何杯のミルクティが必要か？

- 1 杯では、能力が無くても確率 50 % で当たる
- 2 杯でも、確率 25 % で偶然当たる
- 3 杯でも、確率 12.5 % で偶然当たる
- 4 杯でも、確率 6.25 % で偶然当たる
- 5 杯続けて偶然当たる確率は $1/2$ の 5 乗で 3.125 %
- これくらい低い確率なら、偶然ではないと判断してもいいのでは？ という意味で、5% を「有意水準」と決め、能力がなくても偶然当たる確率 (有意確率) がそれ未満なら「判別能力が無い」帰無仮説を棄却する
- この場合、最低 5 杯は試す必要がある (厳密には、テストするミルクティを、ミルクが先、紅茶が先のどちらにするかの順番も考慮する必要がある)

9.3 実験計画におけるサンプルサイズ的设计原則

- 臨床試験の主要なエンドポイント (評価項目) と統計解析の方法が決まっていなくてはならない。
- 割合を比較する場合なら (1) 有意水準, (2) 検出力, (3) コントロール治療での臨床イベント発生割合, (4) 試験治療のイベント発生割合がコントロール治療よりどれくらい小さければ臨床的に意義があると考えられるか、その最小の値

- 例えば有意水準が片側 5%，検出力が 80%，コントロール治療でのイベント発生が 30%，試験治療により期待できるイベント発生が 15%の場合なら，各群 95 人となる。ちなみに R では，

```
power.prop.test(p1=0.15, p2=0.3, sig.level=0.05,  
power=0.8, alternative="one.sided")
```

結果が得られる。

9.4 試験配置法

得られた結果の統計解析についてが，分散分析を用いることが多い。

平行群間比較試験 もっとも単純なデザイン。インフォームドコンセントが得られた適格な患者がランダムに割り付けられ，いくつかの治療のうちの 1 つだけを受ける。得られた結果の統計解析は 2 群なら Welch の t 検定，3 群以上なら一元配置分散分析になる。

乱塊法 すべての処理組み合わせの実験を 1 回ずつ集めたもので 1 つのブロックを形成。ブロック数が反復数になる

分割区法 広い区画で 1 次因子，その区画ごとの細かい条件を 2 次因子とする分割

ラテン方格法 効果を調整したい要因の要因の組み合わせをクロス表にした場合に，行と列のどれをとっても 1~5 の数字が一度だけ出現するように割り付け，その数字を比較したい処理の番号とする

9.5 毒性試験

- 化学物質などについて生物に好ましくない作用の有無またはその強さの程度を調べるための試験
- 試験は、評価する毒性の項目（一般毒性、特殊毒性）、使う生物の種類（哺乳動物、魚など）と形態（全体、組織、細胞など）、曝露経路（経口、吸入、経皮など）、曝露期間（長期、短期など）によって様々な種類がある。目的によって、適切な試験方法を選定する必要がある
- とくに、用量反応関係（量-反応関係ともいう。dose-response relationship）については多くの方法が開発されてきた

9.6 臨床試験

- ヒトに対する実験（侵襲あり）を臨床試験という
- 新薬とか新しい治療法は，モデル動物で効果があるだけではダメで，どうしてもヒトに効くか，ヒトに有害作用がないかどうかを確認する必要がある，臨床試験は必須

- 無駄になってはいけないので、科学的かつ倫理的に考え抜かれた計画に従って行われねばならない。様々なガイドラインがあって、それに沿って計画する必要がある。
- きちんと計画しない臨床試験は許されない（後になってから、サンプルサイズが不十分だったので有意差が見られませんでした、ではダメ）

9.6.1 臨床試験の4段階

- 前段階（非臨床試験）：細胞，組織，動物を使った実験（主として上述の毒性試験）＝安全性を確かめる
- 第I相試験：健康な成人のボランティアを対象として，薬物動態や最大許容量を調べる
- 第II相試験：比較的少数の患者を対象として，有効性，安全性，用量反応関係を調べる
- 第III相試験：数百から数千の患者を対象として，「薬の候補」の有効性を，科学的に検証する目的で行う。通常は **RCT (Randomized Controlled Trial)** になる。
- 第IV相試験：市販後に大勢の患者が実際に服用した結果，新薬がどのような特徴をもっているか，副作用はでないかを調べる（市販後臨床試験）

9.6.2 臨床試験における倫理的要求

- ヘルシンキ宣言や，それに基づいて各国が出している倫理指針が大原則。しかしこれだけでは具体的にどうすればいいのか曖昧。
- 具体的には，日米欧による International Conference on Harmonization (ICH) という会議により，いくつものガイドラインが公表されている。
- 医薬品の臨床試験の実施の基準に関する省令 (GCP)
- 臨床試験のための統計的原則について（統計ガイドライン）
- 臨床試験における対照群の選択とそれに関連する諸問題（対照群に関するガイドライン）

9.6.3 臨床試験の手順

試験実施計画書の作成 実験なので当然。

臨床研究倫理審査 IRB (Institutional Review Board) による倫理審査により研究実施について承認を受ける必要がある

試験の実施 試験実施計画書に沿って実施するのは当然。ただし，有害作用に苦しむ患者が発生した場合，同じ治療を続けることは倫理的に正しくないので，計画書からの逸脱が正当化される（省令 GCP 第 46 条）

データ解析 薬の候補を使う、使わないはランダムに割り付けられるが、倫理的問題を含めて割り付けが守られない（服薬コンプライアンスが悪い）場合があるのが問題。その場合の扱いとしては、実際の服薬にかかわらず割り付け通りに解析する ITT (intention to treat / intent to treat) にするのが原則¹。

9.6.4 ITT の理屈

以下、ITT の理屈を佐藤俊哉、松岡浄 (2001)²に基づいて説明する。臨床試験で A か B の 2 つの処理をランダムに割り付けるとして、実際には割り付けが 100% 守られるとは限らない。結果として以下 4 群が生じる。

Aa 割付けが A, 実際に使ったのが A

Ba 割付けが B, 実際に使ったのが A

Ab 割付けが A, 実際に使ったのが B

Bb 割付けが B, 実際に使ったのが B

全員が Aa または Bb なら問題ないが、実際は Ba や Ab の人が生じる。誰と誰を比べればいい？

- **計画書を守った人たちだけを比べる** Aa を A 群, Bb を B 群とする。これは、治療効果が出たり副作用が出たりした人が減るという選択バイアスがかかるのでダメ。
- **実際に使われた処理で比べる** Aa + Ba を A 群, Ab + Bb を B 群とする。計画書を守っていないのでダメ。Ab を Bb と一緒に B 群に入れると、A が効かなかった人が B 群に入る傾向がでて、やはり選択バイアスがかかる。
- **実際の使用は無視して、割り付け通りに比べる** Aa + Ab を A 群, Ba + Bb を B 群とする。これが正しい。なんで？

なぜ割り付け通りに比べる ITT が正しいのか？

- 割付けを守らなかった Ab と Ba には理由がある⇒中には飲み忘れや交通事故にあって服薬を続けられなくなったなど、治療効果と関係ない場合もあるが、多くの場合ランダムでない（治療が効かないから止めた、治療が効いて良くなったから止めた、副作用が出た、等）。
- A が新薬, B がプラセボだとして、
 - 対立仮説「A が有効」の下では、Aa+Ba を A 群とすると治療効果を良く見せる。Aa+Ab を A 群とすると治療効果を薄める。

¹ただし、ランダム割り付け後に実験参加不適格であることが判明したとか、1 度も薬を飲まなかったとか、ランダム割り付け後のデータが一切ない人については、解析から除外してもいい場合があり、その場合は、残りの「最大の解析対象集団」について、割付通りに解析する。

²<http://www.kbs.med.kyoto-u.ac.jp/01Sep15.pdf>

- 帰無仮説「A は効果なし」の下では、Aa+Ba を A 群とすると治療効果を良く見せるが、Aa+Ab を A 群とすると、正しく「効果なし」と判定できる。
- 計画書を守った人だけ解析したり実際に受けた治療で群分けして解析すると、必ずバイアスがかかるが、割付け通りなら、帰無仮説の下では正しい解析ができるので、**第一種の過誤を守れる**。

言い換えると、ITT をすると差が薄まる場合があるけれども、差が無い場合は正しく解析できる。本当は効かない薬が効いたと判定してしまうよりは、この試験では統計学的に有意な差は検出できなかったという方がマシ。両方やって、一致した結果が得られれば OK。違うときは ITT の方がよさそうというのも一つの判断。

9.6.5 欠損値はどうするか

- ITT の考え方から、最大の解析対象集団を解析するとして、欠損値はどうしたらいいのか？
- 欠損の種類により異なる
 - 完全にランダムな欠損なら検出力が落ちるだけでバイアスは問題にならない
 - 欠損値のある変数についてランダムな欠損で、他の変数についてランダムでない場合は、多重代入法 (Multiple Imputation) の利用を考慮する³。
 - ランダムでない欠損はいかんともしがたい

9.6.6 治療効果の判定指標（その意味）

- 相対リスク減少率（1 - リスク比）
- 絶対リスク減少率（リスク差=超過危険=寄与危険）
- 要治療数 (NNT = 絶対リスク減少率の逆数)

³理屈は難しいが、R では mice などのパッケージで可能。伝統的に単純代入法として用いられてきた、最悪値の代入、直前値の代入、グループ間の差が縮まる値を代入、などは非推奨。

(例) コレステロール低下薬の効果

5年間プラバスタチンを服用した1000人のうち32人が冠動脈疾患で死亡、偽薬を飲んで1000人のうち41人が死亡。新聞報道は、「プラバスタチンを飲むと死亡リスクが22%低下した」
一般市民の多くは、プラバスタチンを飲むと1000人の高コレステロール血症患者のうち220人が心臓発作を免れると誤解したが、それは正しくない。

(Quiz) 3つの判定指標を計算してみると？

1000人中32人の死亡と1000人中41人の死亡の比較という点は同じ。

相対リスク減少率 $1 - (32/1000)/(41/1000) = 1 - 32/41 = 9/41 = 0.2195\dots$

*新聞報道の「22%」はこれだった。

絶対リスク減少率 $(41/1000) - (32/1000) = 9/1000 = 0.009$

*この値を使うなら、「0.9%の死亡を防げる」あるいは、「1000人服用したとき9人の死亡を防げる」といえる⇒効果は劇的に見えないが誤解も少ない

NNT $1/(9/1000) = 1000/9 = 111.11\dots$

*この値を使うなら、「111人服用して初めて1人の死亡を防げる」といえる⇒より誤解されにくい

参考：95%信頼区間の計算

- リスク比の信頼区間は、

$$\exp(\ln(32/41) \pm 1.96 \times \sqrt{(1/32 - 1/1000 + 1/41 - 1/1000)})$$

から、(0.495, 1.223)となる。1から引いて相対リスク減少率を考えると（通常は1を引いて相対リスク増加率を考えるが、この場合は裏返しの方がわかりやすいのでこうする）、(-0.223, 0.505)なので⁴、期待値は0.22かもしれないが0の可能性も十分にある。

- 絶対リスク減少率の信頼区間は、

$$0.009 \pm 1.96 \sqrt{\frac{32 \times (1000 - 32)}{(1000^3)} + \frac{41 \times (1000 - 41)}{(1000^3)}}$$

計算すると(-0.008, 0.255)となり⁵、これも0を含んでいる。

9.6.7 新薬の臨床試験における作業仮説

それまでに標準的な治療法が確立していない場合 プラセボと新薬によるRCTを実施する。新薬投与がプラセボ投与より有効かつ副作用のデメリットを上回ることが必要。

それまでに標準的な治療法がある場合 従来薬と新薬によるRCTを実施する。通常は、新薬は従来薬より有効でなくてはならない。例外として、従来薬の

⁴R の fmsb パッケージを使えば、riskratio(32, 41, 1000, 1000) と打てば計算できる。

⁵R の fmsb パッケージを使えば、riskdifference(41, 32, 1000, 1000) と打てば計算できる。

副作用が大きいとき，新薬の副作用が小さければ，有効性は同等でいいので，「非劣性試験」を行う。他に「非劣性」でいいのは，ジェネリック医薬品の開発の場合。

Chapter 10

スクリーニング

10.1 スクリーニング (Screening) とは？

スクリーニング

集団を対象に、「迅速に実施可能な検査，手技を用いて，無自覚の疾病または障害を暫定的に識別すること」

スクリーンとは元々「篩に掛ける」ことを意味する。疫学や公衆衛生学でスクリーニングといえば，一般には健診や集団検診のこと。

10.1.1 スクリーニングの目的

究極の目的は，集団全体としての死亡率を下げる（普通は発生率は不変）。

主目的 疾病の早期発見，早期治療＝二次予防

他の目的 ある疾病のハイリスク群の検出＝一次予防，集団の特性の把握「集団診断」等

事例：米国乳がんスクリーニング

- 複数の機関による矛盾する勧奨
- 知見も食い違っている
- 有病割合の低さと偽陽性の多さによる不利益
- 2009年11月のUS PSTF (preventive services task force) 勧告を巡る論争
 - － 勧告は，40代は定期的マンモグラフィ不要，50-74は2年おきのマンモグラフィ，75歳以上は不要，マンモグラフィと視触診の併用は無意味，自己触診は過剰診断をもたらす
 - － 新聞，ハーヴァード大学放射線科教授，患者グループなど反論
- 根拠に基づく勧告と誤解に基づく反論？

10.2 スクリーニングにおけるバイアス

スクリーニングにおけるバイアス

自己選択バイアス、リードタイムバイアス、レンジスバイアスがあることに注意

10.2.1 自己選択バイアス

検診受診者は健康維持に関心が高いので生活習慣などもよくなり、健康水準が高くなりがち。そのため、検診の効果が過大評価されがち。

10.2.2 リードタイムバイアス

無症状での検出は症状が出るより前（そのギャップをリードタイム (lead time) と呼ぶ）。仮にまったく早期治療の効果がなくて同じステージで死亡しても生存時間はスクリーニングの方が長くみえてしまう。

10.2.3 レンジスバイアス

スクリーニングでは経過の遅い症例が多く見つかりがちなので、患者全体の死亡率よりスクリーニングで見つかった症例の死亡率が低く見える。

10.3 スクリーニング実施の原則

Wilson and Jungner (1968) は、以下すべてを満たすべきであると提唱した（括弧内は原文）。

1. 目的とする疾患が重要な健康問題である (The condition sought should be an important health problem.)¹
2. 早期に発見を行なった場合に、適切な治療法がある (There should be an accepted treatment for patients with recognized disease.)²
3. 陽性者の確定診断の手段、施設がある (Facilities for diagnosis and treatment should be available.)
4. 目的とする疾病に潜伏期あるいは無症状期がある (There should be a recognizable latent or early symptomatic stage.)
5. 目的とする疾病に対する適切なスクリーニング検査法がある (There should be a suitable test or examination.)³

¹ 「重要な健康問題」が意味するのは、その疾患の頻度が高いことや (eg. がん検診)、緊急な対処が必要なこと (eg. 新生児代謝異常検査) である。

² 治療法がないと「負のラベリング効果」になることがあるため、スクリーニングはしない。

³ 「適切な」は、費用や判定に要する時間も含む。

6. 検査方法が集団に対して適用可能で、受け入れやすい (The test should be acceptable to the population.)
7. 目的とする疾病の自然史がわかっている (The natural history of the condition, including development from latent to declared disease, should be adequately understood.)
8. 患者として扱われるべき人についての政策的合意が存在する (There should be an agreed policy on whom to treat as patients.)⁴
9. スクリーニング事業全体としての費用 - 便益が成立する (The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.)
10. 患者検出は継続して行われる定期検査にするべきで、「全員を一度だけ」対象とする計画ではいけない (Case-finding should be a continuing process and not a "once and for all" project.)⁵

10.3.1 現代における新しいスクリーニング基準

Wilson and Jungner の報告書から 40 年を経て、ヒトゲノム解読が完了した現代に、スクリーニングをすべき基準を再度見直そうとした論文が、Andermann らにより最近多数 publish されている。Andermann *et al.* (2008) は、この 40 年に出現した新しい基準として以下のポイントを挙げている。

1. スクリーニングプログラムは、需要の認識に対応して行われるべきである。
2. スクリーニングの目的が最初の時点で定義されているべきである。
3. 明確に定義された対象集団が存在すべきである。
4. スクリーニングプログラムの有効性について、科学的な証拠が存在すべきである。
5. スクリーニングプログラムは、教育、検査、臨床サービス、プログラム管理を統合すべきである。
6. スクリーニングの潜在的なリスクを最小化するメカニズムを含む質の保証が存在すべきである。

⁴この項、境界例をどう扱うかが予め決まってい適切な処遇が受けられること、という主旨であるため、これまで「患者、要観察者に対する追跡 follow-up システムが確立している」としていたが、直訳にした。

⁵一度だけのスクリーニングでは、集団中の限られた割合の人（健康に関心があるリスクが少ない人）しかカバーできないとか、そのとき気になる症状をもっている人は受診するかもしれないが、それから将来の疾病発生リスクを予測することはできないので、地道な健康教育などを通じて定期的な健診が保健医療サービスの一つであることを周知することによって、本当にスクリーニングを受けて疾病を早期発見することが有用なハイリスク層を検査することができるので、定期的に継続すべきとしている。この項は、これまで「スクリーニングの意味、内容が受診者に周知されている」とし、受診者が過度な期待をしないことにより、不要な争いを避けることができる、という効果を書いていたが、本来の意図とずれていたため 2020 年 8 月 25 日に改訂した。

7. スクリーニングプログラムは、インフォームドチョイス（十分な情報提供を伴った上での選択）、信頼性、患者の自律性の尊重を確証すべきである。
8. スクリーニングプログラムは、対象集団全体に対して公平なアクセスを推進すべきである。
9. 実施後の結果から、スクリーニングプログラムの評価が計画されるべきである。
10. スクリーニングの総体としての利益は、有害事象（による損失）を上回っているべきである。

これらの点を踏まえ、Andermannらは、遺伝的スクリーニング (genetic screening) の実施についての考察を展開している。Andermann *et al.* (2009) はガイドラインであり、Andermann and Blancquaert (2010) はプライマリケア専門家へのメッセージである。このことには、倫理的な問題が大きくかかわってくるのは自明であろう。玉井 (2006) が指摘するように、遺伝情報には以下3つの特殊性がある。

- 不変性・予見性：生涯変わることがなく、ときに将来の発症を予見しうる。
- 個人特異性：DNA の配列そのもので個人を特定しうる。
- 家系内共有性：血縁者の遺伝的リスクまでもが否応なく明らかにされうる。

とくに家系内共有性があることにより、家族が受けて遺伝子疾患をもっていることが明らかになった場合、自分が遺伝診断を受けなくてもかなりの確率で同じ疾患を発症する可能性があることがわかってしまうため、自己決定の意味が問い直されることになる。玉井 (2006) は、出生前診断⁶のために遺伝外来を訪れたある女性の語りを紹介している。

重い障害をもつ子どもがいる。次は健康な子がほしい。でも、上の子の存在を否定しているような気がして、出生前診断を受けることに罪悪感を覚えてしまう。似たような状況にある人はみんな受けてるじゃないか、と言いつつ考えている自分がいやだ。でも、もうひとり同じ障害の子が生まれたら、育てていけるだろうか。

Andermann and Blancquaert (2010) の主要なメッセージは、『**遺伝的スクリーニングは無症状の個人からなる特定の集団に系統的に提供される公衆衛生プログラムの一種であり、ハイリスクであると同定された人々に予防、早期治療、あるいは再生産上のオプションを提供することを目的としている。通常の臨床ケアに比べてスクリーニングによって付加される利益を確実にし、意図しない有害事**

⁶母体内で育ちつつある胎児の健康状態を、生まれる前に把握するために行われる種々の検査と、それらの検査結果に基づく診断行為の総称。広義の出生前診断は、妊婦や胎児の健康管理に有用であり、適切な分娩方法・分娩施設を選択することや、出生後すみやかに適切な医療的処置を行うためにも役立つものであり、妊娠の全期間を通して必要に応じて行われる。一方、狭義の出生前診断とは、胎児の疾患を積極的に発見し、妊娠を継続するか否かの判断材料とするために、中絶可能な時期（妊娠 22 週未満）に検査結果が出ることを前提として行われる。後者が優生思想につながるものが倫理的に問題となる。胎児にとっては、自分の存在が抹殺されるかどうかを決めるのに使われる情報なのに、その胎児の生体情報にアクセスする権利が親にあるのか？

象（不必要な不安とかレッテル張り）を予防できるかどうかは、スクリーニングプログラムのデザインと実施（対象者の募集方法、提供される教育とカウンセリング、スクリーニングのタイミング、検査的中率、利用できる介入、監視メカニズムとセーフガードの存在を含む）にかかっている。経済的な関心だけでスクリーニングプログラムを計画すると、プログラムの有効性、受容可能性、適切性が示される前にスクリーニングを導入し拡大してしまうような市場原理に基づいたアプローチにつながるかもしれないという心配がだんだん大きくなってきている。どのような医学的介入についても、遺伝的スクリーニングは有害なことより多くのよいことをすべきという道徳的要請がある。これは、個人あるいは家族の視点からだけではなく、対象集団や社会全体としての視点からみてもいえる』ということである。これを踏まえて彼らは、『プライマリケアの専門家は、患者に対し、新しい遺伝技術あるいはゲノム技術の利点とリスクについて十分に知らせることと、それによって彼らがより十分なインフォームドチョイスができるように力を与えることによって、彼らが急速に変化しつつある多様な遺伝的スクリーニングサービスとどのようにつきあっていけばいいか助言する役割を果たしている』と結論している。遺伝的スクリーニングはマススクリーニングから遺伝子診断までの連続体であり、そのどの場面でも、患者と直接接するプライマリケア医の役割は重要である。

10.4 スクリーニング方法の評価

スクリーニングの性能評価指標

有効性（感度、特異度）と信頼性（再現性）の他、迅速性、簡便で安価であることが求められる

10.4.1 有効性の指標

有効性の指標としての感度と特異度を計算するには、**確定診断（臨床診断あるいは既に確立している Gold Standard の検査法による）**がついている疾病をもつ人と健康な人の両方について検査を行い、検査の結果が陽性か陰性かを明らかにする必要がある。感度 (sensitivity) は positive in disease, 特異度 (specificity) は negative in health と覚えるといい。2 × 2 の分割表を使って説明すると以下の通り。

	疾病	健康
検査陽性	a 人	b 人
検査陰性	c 人	d 人

感度 $\frac{a}{(a+c)}$

特異度 $\frac{d}{(b+d)}$

陽性尤度比 $\frac{(a/(a+c))}{(b/(b+d))}$ 感度を (1 - 特異度) で割った値。

陰性尤度比 $\frac{(d/(b+d))}{(c/(a+c))}$ 特異度を (1 - 感度) で割った値。

これらと紛らわしいが、実際にスクリーニングを行った場合に得られる判定の的中度は、**対象となる集団における有病割合に依存する**ので、混同してはならない。前述のものと同一の 2×2 の分割表を使って説明すると（ただし、この場合は実際のスクリーニングなので、事前には疾病の有無はわからず、事後的に精密検査等で疾病の有無を確定して得られる数値と考えて）、以下の通り。

陽性反応的中度 (positive predictive value) $\frac{a}{(a+b)}$

陰性反応的中度 (negative predictive value) $\frac{d}{(c+d)}$

稀な疾病（有病割合がきわめて低い）では $(a+c) \ll (b+d)$ なので、**感度と特異度がともに高くても陽性反応的中度が低くなる**ことがある。橋本愛主演のドラマ『ハードナッツ!』で、伴田刑事が「1万人に1人がかかる難病で、検査精度は10000人が受けに来たら、9990人が正しい判定ということ」の陽性判定が出て落ち込んでいるところ、ヒロインくるみが「1万人が検査を受けたとすると（…中略…）10人には間違っただ判定が下る。ということは、本当に病気なのは1万人にひとりのはずなのに、病気と判定される人が10人くらい出てきてしまうってことなんです。だから、伴田さんの病気の確率は、99.9%じゃなくて、10人のうちのひとり、せいぜい10%ってことなんです」と、結局は、有病割合が低いから陽性反応的中度はせいぜい10%くらいだろうという話に帰着する。厳密に言えば、 $a/(a+b) = a/(2a+9)$ 、かつこの関数は $a = -9/2$ の点を除けば連続で単調増加なので、 $a/(a+b)$ が取れる値は、 $0 \leq a \leq 1$ では $a = 1$ のとき最大になり、 $1/11$ と、「せいぜい10%」はまったく正しい。

10.4.2 確定診断に用いる検査方法自体の感度・特異度？

2020年にパンデミックを起こしている COVID-19（新型コロナウイルス感染症）の確定診断には RT-PCR 検査が用いられるが、この検査結果がどれくらい信頼できるのかということが問題になり、ある程度疫学を知っているはずの医師などでも感度、特異度、陽性反応的中度について混乱が見られた。

既に述べた通り、感度は、既に疾病ありと確定診断がついている人を検査したときに正しく陽性と判定できる割合であり、特異度は、疾病なしとわかっている人を検査したときに正しく陰性と判定できる割合であった。ところが、COVID-19で疾病ありと確定診断を付ける方法自体が RT-PCR での検査陽性なので、検査前に疾病の有無を知ることができない。そのため、感度も特異度も計算できない。無理矢理感度や特異度を計算すると100%になってしまう。

症例定義の要件に含まれる検査項目について感度や特異度を考えることは原理的にはおかしい。しかし、2020年8月現在、RT-PCRによる COVID-19の原因ウイルスである SARS-CoV-2 検出の性能は、感度が高くても80%程度、特異度は99.99%以上と言われている。これはどうやって計算しているのかといえば、時間をおいて何度か検査すると、最初陰性だった人が陽性になる場合が多々あることから、本当は感染していたのに最初はそれを検出できなかったという推論に基づいている。しかし、1回目に測ったときは本当に感染していなくて、後で感染したのかもしれない、という可能性を完全に排除することはできない。この可能性を聞き取りからできる限り排除して計算したことにより、2020年2月末頃に言われていた値よりは感度も特異度も良いことがわかったが、厳密な感度や特異度は、後付けでは計算できない。

複数の検査の組み合わせによって、感度か特異度を上げることができる。ただしトレードオフである。例えば、2つの検査がともに陽性だった場合のみ陽性と判定する、としたら、特異度は上がるが感度は下がる。逆に、どちらか1つでも陽性なら陽性とする、としたら、感度は上がるが特異度は下がる。現状、SARS-CoV-2のRNA検出のプローブは2つ使われていて、1つはSARS-CoV-2に特異的な配列で、もう1つは他のウイルスでも持っている可能性が多少はあるものらしい（この辺、ちゃんと文献を読み込んでいないので間違っているかも。ただ、この流行初期にHIVのRNAが入っているとした誤報は、HIVだけではなく、コロナウイルスにも元々あって不思議はない配列を検出のターゲットにしてしまったことによるので、SARS-CoV-2なら共通して持っているが他のウイルスは持っていない、という配列を決めるのは、そう簡単なことではないようだ）。少なくともいくつかの論文では、ともに検出されたときのみ陽性と判定されていたので、その分、特異度は上がるが、感度は低くなる。確定診断法である以上、特異度が高いことが決定的に重要なのは確かで、この方針は合理的である。

もっと難しいのは、すべての検査には検出限界があるということだ。狂牛病が流行したとき、日本の農水省は、人々の「安心」を重視して全頭検査を実施したが、生後1年未満の仔牛ではプリオン濃度が検出限界以下なので、検査しても検出される可能性がなく、生後1年以上の牛だけ検査すれば十分だという批判があった。あの全頭検査には、確かに安心を与える以上の意味はなかった。ただ、検出限界以下のプリオンしか含まれていない肉であれば、nvCJDを起こす可能性はほとんどないと考えられたので、一様に検査することで「安心」は得られた。COVID-19の場合、咽頭や鼻腔からのスワブ（ウイルス濃度が低ければ、体内にウイルスがいても偶々スワブにウイルスが付かない可能性がある）、あるいは唾液中（感染初期はあまりウイルスが含まれていない可能性が高い）の濃度が検出限界以下（おそらく無症状の人に多いと思われる）であっても、その人に感染力がないとは限らない。微量なままでも、マスク無しでの会話で飛ぶ唾液などに混じっていたウイルスが、飛んだ先の人の体内に入って増殖してしまうかもしれない。従って、検査方法としては、検出系だけではなく、サンプリングの方法と検出限界を含めて評価することが非常に重要であることは知っておいた方がよい。

10.4.3 信頼性

スクリーニングの信頼性は、検査再検査信頼性 (test-retest reliability)、施設間差 (inter-institute difference) が小さいこと、測定者間差 (inter-rater difference) が小さいことなどを指す。

test-retest reliability は、同じ対象者に同じ検査をしたときに結果が一致すること。 κ 係数 = 実際の一致率と偶然の一致率の差を、1 と偶然の一致率の差で割った値。完全一致のとき 1、偶然の一致と同じとき 0、それ以下で負となる。

10.5 連続量の測定における ROC 分析

- ROC は、Receiver Operating Characteristic の略語。普通、ROC 曲線という形で使われる
- 日本語では受診者動作特性曲線という訳語がついている教科書と、受信者動作特性曲線という訳語がついている教科書が並立しているが、ROC が何の

略であるかを明示して「ROC 曲線」だけを掲載している本も増えてきた。

- 日本疫学会（編）「疫学 基礎から学ぶために」南江堂，能登洋「日常診療にすぐに使える臨床統計学」羊土社などが「受診者」
- 鈴木・久道（編）「シンプル衛生公衆衛生学 2006」南江堂，日本疫学会（訳）「疫学辞典 第3版」日本公衆衛生協会，フレッチャー RH，フレッチャー SW，ワグナー EH，福井次矢（監訳）「臨床疫学」メディカルサイエンスインターナショナルなどが「受信者」
- 稲葉・野崎（編）「新簡明衛生公衆衛生 改訂4版」南山堂，丹後俊郎「メタ・アナリシス入門」朝倉書店などは、「ROC 曲線」だけ

10.5.1 ROC 曲線とは

- 1つのスクリーニング方法について陽性・陰性の基準値を最小値から最大値まで段階的に変えると，偽陽性率（ $=1 - \text{特異度}$ ）も感度も変化
- 基準値=最小値で感度も偽陽性率も 1，基準値=最大値で感度も偽陽性率も 0
- 偽陽性率を横軸，感度を縦軸にとって線で結ぶと，基準値の変化に対応する「ROC 曲線」になる。
- 左上 (0,1) 近くを通るほどスクリーニング有効性が高い（実際は費用など他の条件も考慮）
- 曲線上で一番左上に近い点が最適なカットオフ値といえる
- 最適カットオフ値の探索にも使えるし複数のスクリーニング方法の優劣の判定にも使える
- ROC 曲線が左上近くを通ることと AUC（Area under curve; 曲線下面積）が大きいことは同値

10.5.2 ROC 曲線を描く例

10 人の対象者について，うつ質問紙による得点と臨床診断（架空のもの）が下表のように得られているとする。

対象者	得点	臨床診断
1	20	うつ
5	22	うつ
6	28	うつ
2	13	健康
3	19	健康
4	21	健康
7	11	健康
8	25	健康
9	16	健康
10	19	健康

仮にカットオフ値が 18（18 以上を陽性と判定）だとすると、 2×2 分割表は、

	うつ	健康
陽性	3	4
陰性	0	3

となる。このときの感度は $3/3 = 1$ 、特異度は $3/7 = 0.429$ 、偽陽性率は 0.571。カットオフ値を小さくすれば感度が上がって特異度が下がり、逆にカットオフ値を大きくすれば感度が下がって特異度が上がる。R では下記のコードで実行できる。

```
x <- c(20,22,28,13,19,21,11,25,16,19)
y <- c(1,1,1,0,0,0,0,0,0,0)
# Epi ライブラリを使う場合
require(Epi)
ROC(x,y,plot="ROC")
# fmsb ライブラリを使う場合
require(fmsb)
res <- roc(x,y)
print(res)
plot(res)
```

最適カットオフ値は 20 で、そのとき感度 1、特異度 0.714 とわかる。AUC は 0.857 である。

10.6 母子保健分野の代表的なスクリーニング

10.6.1 新生児マススクリーニング検査

2003 年遺伝医学関連 10 学会発表『遺伝学的検査に関するガイドライン』の指針には、下記のように定められている。

1. 新生児の先天性疾患の早期診断・早期治療による発病率・死亡率の低下を目的
2. 医師は検査の意義を積極的に両親に説明し同意を得て検査実施に努め、必要に応じて遺伝カウンセリングを考慮する

発達障害については、1 歳 6 カ月児健診でも 3 歳児健診でも 84~85% の自治体で実施されている。ただし、そのうち 8 割は問診の中で聞き取っているだけ。13~14% の自治体は発達障害の疑いのある児に特別な検診を用意している。

先天性代謝異常等については、ハイリスクグループ検出と早期治療が目的で実施される。

- 全新生児中保護者が検査を希望した者（実際はほぼ全数）対象
- 生後 1 週間以内、足底動脈から採血、ろ紙に吸着・乾燥させ、センターに送付して検査
- フェニルケトン尿症、楓糖尿症、ホモシスチン尿症、ガラクトース尿症、先天性副腎過形成症、クレチン症を検出

- 公費医療により迅速対処
- 平成 20 年度 108 万人強の新生児に対し、約 116 万件の検査（2kg 以下低体重児は再検査するため）、先天性代謝異常 125 例、先天性副腎過形成 49 例、クレチン症 606 例を検出
- 神経芽細胞腫については、1984 年～2003 年、生後 6～7ヶ月児対象、尿検査でマス・スクリーニングがなされる。神経芽細胞腫は小児がんの一種で、カテコラミンを産生するので、中間代謝産物の VMA や HVA を測定して検出される。1 歳未満で見つかりと予後が良く、1 歳以降で見つかりと予後が悪いことがスクリーニング実施の根拠だったが、ドイツとカナダで行われた介入研究の結果、2002 年に有意な死亡率減少効果がなかったという知見があり、スクリーニング中止された。検査開始後累積罹患率が 2 倍になっていたことから、過剰診断が疑われた。スクリーニングで検出された中に自然退縮例が多いことも知られている。手術例の 10%、化学療法例の 5%に治療による合併症がみられた。ただし、1995～2000 年出生児約 716 万人対象の前向きコホート研究の 2001 年中間報告で、受診群の死亡が未受診群の死亡より有意に低かったという有効性が示されたが、検診事業中止により研究も中止されたので、結論は出ないまま。

聴覚スクリーニングについては、早期発見、早期治療が有効であることが示されている。乳幼児健診での聴力検査は 3 歳時（母子保健法施行規則）だが、1990 年代後半に耳音響放射 (OAE) や自動 ABR（聴性脳幹反応）が開発され、米国では 2005 年には約 93%の新生児が受検した。日本は自治体により受検割合に大差。入院中に平均 60%程度実施されている。

10.7 成人期以降の代表的なスクリーニング

10.7.1 特定健診：高齢者医療確保法による

- 40 歳～74 歳を対象
- メタボリック・シンドロームを検出する目的
- 内科系 7 学会合同提案によるスクリーニング基準値（必須項目として男性 85cm 以上、女性 90cm 以上の腹囲または内臓脂肪断面積 100 平方 cm 以上＋以下のうち 2 項目以上：TG150mg/dL 以上、HDL コレステロール 40mg/dL 未満、SBP/DBP が 130/85mmHg 以上、空腹時血糖 110mg/dL 以上）の信頼性・妥当性には反対意見多数あり（まだ変わっていない）
- 中年以降の男性の半数が予備軍になってしまうのでハイリスクグループ検出という意味は無い
- 特定保健指導の有効性にも疑問があるのでポピュレーション・アプローチとしても疑問がもたれている

10.7.2 がん検診：がん対策基本法による

- 部位別にさまざまな検査がある
- 2007年の受検割合は2～3割
- 国立がんセンター「がん検診ガイドライン」⁷

子宮頸がん 細胞診のみ、相応な根拠あり（HPV検査は子宮頸がん死亡率減少効果の有無を判断する証拠が不十分）

胃がん X線検査と内視鏡検査は、死亡減少効果を示す相応な根拠あり

大腸がん 便潜血検査は十分な根拠あり。大腸内視鏡検査は死亡減少効果を示す根拠はあるが無視できない不利益があるため対策型検診は非推奨

肺がん 胸部X線（高危険群に対しては喀痰細胞診併用）は相応な根拠あり

乳がん 40～74歳でマンモグラフィ単独で死亡率減少効果を示す相応な根拠あり、偽陽性、過剰診断、放射線誘発乳がんの発症の可能性という不利益もあるが対策型検診推奨。40～64歳については視触診併用も同様の効果と不利益を示す。

10.7.3 その他の検診

脳動脈瘤・脳梗塞 人間ドックで普通行われる神経学的検査や眼底検査は信頼性△。精密検査あるいは脳ドックでのCTは信頼性○，MRIは信頼性◎

肺結核 胸部X線は老人保健法の基本健診にはあったが高齢者医療確保法の特定健診では消滅

狭心症・心筋梗塞 心電図は特定健診にもあり信頼性は○

高脂血症 基本健診ではTGと総コレステロールだけだったが特定健診になってHDLとLDL（信頼性◎）が入った

貧血 ヘモグロビンが信頼性◎。特定健診に含まれる

肝機能 AST，ALT，γ-GTPは特定健診にも含まれ信頼性も◎

糖尿病 特定健診に含まれる空腹時血糖とHbA1cは信頼性◎，眼底検査による糖尿病性網膜症検出は信頼性○

尿路結石 人間ドックにある腎超音波は信頼性○。尿たんぱくは信頼性△

痛風 人間ドックにある血清尿酸値は痛風の前臨床症状である高尿酸血症の信頼性が◎な指標

⁷<http://canscreen.ncc.go.jp/guideline/matome.html>

10.8 スクリーニングの参考文献

- Andermann A, Blancquaert I, Beauchamp S, Déry V (2008) Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization*, 86(4): 317-319.⁸
- Andermann A, Blancquaert I, Beauchamp S, Costea I (2011) Guiding policy decisions for genetic screening: developing a systematic and transparent approach. *Public Health Genomics*, 14: 9-16⁹
- Andermann A, Blancquaert I (2010) Genetic screening: A primer for primary care. *Canadian Family Physician*, 56: 333-339.
- ゲルト・ギーゲレンツァー (著), 吉田利子 (訳) (2010) 『リスク・リテラシーが身につく統計的思考法』ハヤカワ文庫 (注: この本は疫学でも紹介した)
- 李啓光 (2010) 『続・アメリカ医療の光と影 (第 168 回～第 170 回): 乳癌検診をめぐる大論争①～③』, 週刊医学界新聞, 医学書院, No.2867 (2010年2月15日), No.2869 (2010年3月1日), No.2871 (2010年3月15日)
- 玉井真理子 (2006) 『遺伝医療とこころのケア: 臨床心理士として』NHK ブックス
- 坪野吉孝 (2004) 『「がん」は予防できる』講談社+α新書
- U.S. Preventive Services Task Force (2009) Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*, 151(10): 716-726.¹⁰
- 鷺崎 誠 (2003) 『健康診断・人間ドック「気になる」疑問』角川 one テーマ 21
- Wilkinson MJB, Barczak P (1988) Psychiatric screening in general practice: comparison of the general health questionnaire and the hospital anxiety depression scale. *Journal of the Royal College of General Practitioners*, 38: 311-313.¹¹
- Wilson JMG, Jungner G (1968) *Principles and practice of screening for disease*. World Health Organization, Geneva, 163 pp.¹²

⁸<http://www.who.int/bulletin/volumes/86/4/07-050112.pdf>

⁹この論文は JSTOR に入っているので, 大学内のネットワークからなら全文読める。https://doi.org/10.1159/000272898

¹⁰<http://www.ahrq.gov/clinic/uspstf09/breastcancer/brcanrs.pdf>

¹¹<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1711493/pdf/jroyalcgprac00019-0020.pdf>

¹²http://whqlibdoc.who.int/php/WHO_PHP_34.pdf または <http://www.who.int/bulletin/volumes/86/4/07-050112bp.pdf>

Index

- 5W-Bridge, 8
- case control study, 14
- clinical trial, 16
- cross sectional study, 14
- ecological fallacy, 14
- Ecological Study, 12
- ICD, 9
- longitudinal study, 14
- Natural Experiment, 12
- RCT, 16
- Snow, 12
- 因果関係, 8
- 横断的研究, 14
- オッズ, 22
 - 疾病, 22
 - 曝露, 22
- オッズ比, 14, 28
- 介入研究, 9, 11, 16
- 観察研究, 11
- 患者対照研究, 14
- 記述疫学, 9
- 寄与危険, 28
- クラスターサンプリング, 17
- ケースコホート研究, 15
- 原因, 8
- 健康事象, 8
- 交絡, 14
- 国際疾病分類, 9
- コホート研究, 15
- コホート内症例対照研究, 16
- サンプルサイズ, 19
- 死因简单分類, 10
- 死因分類, 10
- 自然実験, 12
- 死体検案書, 10
- 悉皆調査, 17
- 疾病オッズ, 21
- 疾病分類, 8
- 疾病量, 9, 21
- 死亡診断書, 10
- 縦断的研究, 14
- 集落抽出法, 17
- 症例対照研究, 14
- 生態学的研究, 12
- 生態学的誤謬, 14
- 全数調査, 9
- 相対危険, 28
- 総体危険, 15
- 単純無作為抽出法, 17
- 断面研究, 14
- 地域相関研究, 12
- 治験, 16
- 超過危険, 15, 28
- 曝露オッズ, 21
- 標本抽出法, 9
- 頻度, 8

分析疫学, 9
分布, 8

有病割合, 14, 21

ランダムサンプリング, 17

罹患率, 15, 21, 22

罹患率差, 15, 28

罹患率比, 15, 28

リスク, 15, 21

リスク差, 15, 28

リスク比, 15, 28

臨床試験, 16

累積罹患率, 15, 21