

いろいろなバイアスの例 (5)

- 肺気腫への喫煙の影響を調べるコホート研究
 - 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、喫煙者が来院する时必须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、上記診断率の差はある
 - 「追跡の偏り」による差別的誤分類。情報バイアスの1つ
- 先天異常をもつ子どもを症例。先天異常がない子どもを対照とする症例対照研究で、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、過去について真実に原因を考えて思い出す。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緯を思い出す割合は高い
 - 「母性の思い出しバイアス」による差別的誤分類。情報バイアスの1つ

過去の有名な疫学研究

- Snow J (1813-1858) のコレラの研究 (後述)
- 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 年に大流行したため、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) の脚氣の研究: 1884 年に脚氣の原因が食事の欠陥にあると見当をつけ、脚氣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り、毒素に対する窒素割合を増やしたら、脚氣が減った。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。

久山町研究

- 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 (通常毎年十一斉 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
- 主な結果: 脳出血による死は脳梗塞による死亡の 1.1 倍 (第 1 集団)。脳卒中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率を大幅に減らした介入成果、等

実験計画における Fisher の 3 原則

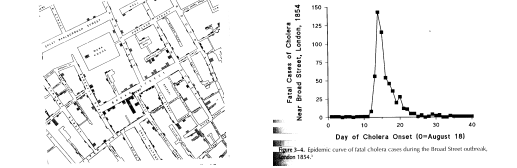
- 反復: 1 つの処理に対して少なくとも 2 回以上の繰り返しが必要
- 無作為化: 実験の順序や空間的などの場所にどの実験群を割り当てるのかを無作為に決める必要がある
- 局所管理: 実験が大規模で、実験全体を無作為化するのが妥当でないとき、実験をある程度細分化してブロックを構成し、ブロック内で処理条件を無作為化し、ブロック内のバックグラウンドが均一になるよう管理すると、系統誤差の一部がブロック間変動として除去できる。

交絡によるバイアス

- 交絡とは、注目している因果関係を歪める第 3 の因子
- 交絡要因の 3 条件 (例: 肥満 ⇒ 高血圧に対する年齢)
 - 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
 - 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)
- 交絡要因を見逃すと変な因果推論をしてしまう (交絡によるバイアスの例) 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と、総死亡率を総人口で割った「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くなってしまふ。年齢構造が交絡要因 (次のスライド)。

Snow によるロンドンのコレラ研究

- 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。
- 1854 年の大流行時、John Snow はコレラ死亡者の発生地図 (spot map) (下図左、出典は I.E.A. [Last JM ed.], 2001) と発症日別のコレラ死者数の度数分布 (下図右、出典は Rothman, 2002) を作成 ⇒ 記述疫学研究。
- 地図からソーホーのコレラ流行が 1 つの共同井戸利用に起因したと推定
- 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&W 社: 4093/266516 と約 5.8 倍) を計算して、S&W 社から水供給を受けている人が危険なことを示した ⇒ 自然実験研究。



INTERSALT 研究

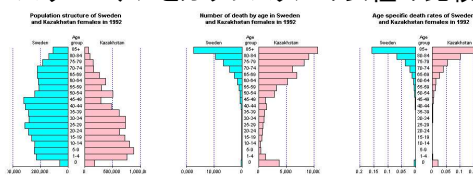
- 32 か国 52 集団 (うち日本 3 集団)、20-59 歳の男女、1 集団 200 例ずつ、1985 年に横断的に実施
- 登録数 10079 人 (男性 5045 例、女性 5034 例)
- 調査項目 血圧 (5 分間の安静のち、ランダムゼロ血圧計を用いて座位で 2 回測定)、24 時間蓄尿 (尿量、ナトリウム、カリウム、マグネシウム、カルシウム)、心拍数、運動状況、飲酒、喫煙、服薬状況、年齢、身長、体重、人種、教育、経産歴 (女性のみ)、閉経状況 (女性のみ)。
- 主な結果: 地域相関で、BMI ∝ 血圧、Na/K 比 ∝ 血圧、塩分摂取量 ∝ 血圧、ヤノマミなど低塩分摂取集団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



実験計画法の発想

- 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、飲むと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
- 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

スウェーデンとカザフスタンの女性の比較



- 水色がスウェーデン、ピンクがカザフスタンの女性のデータ
- 左から、年齢 5 歳階級別人口、年齢階級別死亡率、率
- 年齢階級別死亡率は、どの階級でもカザフスタンが上
- 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 (→ スウェーデンの方が高齢者が多いことにより、地域と死亡率の関係が交絡を受けている)

代表的な疫学研究

- 大規模コホート研究
 - Frammingham Heart Study (フラミンガム研究)
 - 久山町研究 (http://www.epi-c.jp/e001_1_0001.html)
- 国際的な横断研究 + 地域相関研究
 - INTERSALT 研究
- 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

疫学研究倫理指針 → 人を対象とする医学系研究に関する倫理指針

- 2002 年に文部科学省と厚生労働省が合同で発表した指針 (2007 年全面改訂、2008 年 12 月一部改訂、2014 年から臨床研究倫理指針と一体化)。ヘルシンキ宣言やリスボン憲章に依拠している
- 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必要 (侵襲を伴わない場合は不要とされるが、得ておくことを求められる場合が多い)
 - 完全匿名、連絡不能なら記名同意書は不要 (矛盾するので)
- 倫理審査委員会による審査を通らないと研究できない
 - 一例外として既に公開されている二次資料の分析やメタアナリシスなど

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
 - 本当かどうか名だたる学者が大論争
- 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% を有意水準とし、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)

交絡の制御

- 交絡因子によって層別し、別々に解析
 - 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
- 交絡因子の影響を制御
 - 標準化
 - プール化 (マンテルヘンツェルの要約オッズ比などあるが高度な話なので省略)
 - 限定
 - 多変量解析 (ロジスティック回帰分析など)

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
- 疫学的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、子世代コホート (1971 年) 5124 人、第三世代コホート (2005 年) 4095 人
- 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 (子世代コホート) X 線撮影を除く上記項目 + 心エコー、動動脈エコー、ホルター心電図
- <http://www.framinghamheartstudy.org/>



二次資料の解析のため倫理審査不要な研究例

Thyroid Cancer Detection by Ultrasound Among Residents Ages 18 Years and Younger in Fukushima, Japan: 2011 to 2014

Toshihiko Tsuda¹, Akiko Tokimura², Eiji Yamamoto³, and Eungil Seung⁴

Epidemiology, 27(3): 316-322, 2016. [<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4282066/>]

Background: After the Great East Japan Earthquake and Tsunami in March 2011, radioactive elements were released from the Fukushima Daiichi Nuclear Power Plant. Based on prior knowledge, concerns emerged about whether an increased incidence of thyroid cancer among exposed residents would occur as a result.

Methods: After the release, Fukushima Prefecture performed ultrasound thyroid screening on all residents aged 4–17 years. The first round of screening included 296,577 examinees, and a second round began in April 2014. We analyzed the prevalence results from the first and second round up to December 31, 2014, in comparison with the Japanese national incidence and the incidence within a reference area in Fukushima Prefecture.

Results: The highest incidence rate ratio, using a latency period of 4 years, was observed in the central middle district of the prefecture compared with the Japanese annual incidence (incidence rate ratio = 20; 95% confidence interval [CI] = 25, 90). The prevalence of thyroid cancer was 60.9 per million examinees (95% CI = 302, 1000) and the prevalence odds ratio compared with the reference district in Fukushima Prefecture was 16 (95% CI = 10.99, 27.10). In the second screening round, even under the assumption that the rest of examinees were disease free, the incidence rate ratio of 12 has already been observed (95% CI = 5.1, 23).

Area and District (Year)	Prevalence of Thyroid Cancer (per 10 ⁶ CI)	Prevalence Ratio	95% CI
Japan (2011) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2012) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2013) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2014) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2015) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2016) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2017) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2018) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2019) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2020) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2021) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2022) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2023) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2024) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2025) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2026) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2027) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2028) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2029) (total area)	69,208 (52)	1.0	0.81–0.89
Japan (2030) (total area)	69,208 (52)	1.0	0.81–0.89

サンプルサイズの設計原則

- 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。(R で計算する場合 = `power.prop.test(p1=0.15, p2=0.3, sig.level=0.05, power=0.8, alternative="one.sided")` で OK)。

試験配置法

- 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
- 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

Oct. 28, 2019

33

倫理的要求

- ヘルシンキ宣言が大原則。しかしこれだけでは具体的にどうすればいいのが曖昧。
- 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)

Oct. 28, 2019

37

なぜ割り付け通りに比べるか

- 割り付けを守らなかったAbとBaには理由がある
⇒多くはランダムではない(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)。中には飲み忘れや交通事故にあつ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。
- AbやBaを解析から省くと、治療効果が出たり副作用が出たりした人が減る選択バイアスがかかる
- 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

Oct. 28, 2019

41

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)
- (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、偽薬を飲んでた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
- (Quiz)3つの判定指標を計算してみてください

Oct. 28, 2019

45

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

Oct. 28, 2019

34

臨床試験の手順

- 試験実施計画書の作成: 実験なので当然、試験実施計画書に沿った試験の実施も当然。ただし以下の問題あり。
 - 倫理的問題: 有害作用に苦しむ患者に同じ治療を続けられるか? →この場合は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場合があるのが問題。その場合の扱いとして、ITT(intention to treat / intent to treat)は重要。

Oct. 28, 2019

38

Intention-to-treat ではどうか

- Aが新薬、Bがプラセボだと、
 - 対立仮説「Aが有効」の下では
 - Aa+Bb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
 - 帰無仮説「Aは効果なし」の下では
 - Aa+Bb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、正しく「効果なし」判定
- 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
- 割り付け通りなら、帰無仮説の下では正しい解析 ⇒第一種の過誤を守る

Oct. 28, 2019

42

解答

- 1000人中32人の死亡と1000人中41人の死亡の比較
- 相対リスク減少率は、 $1 - (32/1000)/(41/1000) = 1 - 32/41 = 9/41 = 0.2195...$
*新聞報道の「22%」はこれだった。
- 絶対リスク減少率は、 $(41/1000) - (32/1000) = 9/1000$
*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は $1/(9/1000) = 1000/9 = 111.11...$
*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

Oct. 28, 2019

46

臨床試験とは

- 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つて、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)

Oct. 28, 2019

35

ITTの理屈(出典:佐藤先生 pdf)

- 臨床試験データには、AかBの2つの処理をランダムに割りけるとして、
 - 割り付けがA、実際に使ったのがA = Aa
 - 割り付けがB、実際に使ったのがA = Ba
 - 割り付けがA、実際に使ったのがB = Ab
 - 割り付けがB、実際に使ったのがB = Bb
- の4群が生じる。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のか? ……ちよつと考えてみてください

Oct. 28, 2019

39

割り付けが守られない場合はITT

- 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、倫理など様々な理由で割り付けが守られない場合がある
- 守った人だけ(計画書に適合した対象集団)で解析する
 - 選択バイアスがかかる
- 最初の割り付け通りに解析する(ITT)ただし、ランダム割り付け後に実験参加不適合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、割り通りに解析する)
 - 差が薄まるけれども、差が無い場合は正しく解析できる
- 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

Oct. 28, 2019

43

参考: 95%信頼区間の計算

- リスク比の信頼区間は、 $\exp(\ln(32/41) \pm 1.96 * \sqrt{(1/32 - 1/1000 + 1/41 - 1/1000)}) = (0.495, 1.223)$
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、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、 $0.009 \pm 1.96 * \sqrt{(32 * (1000 - 32) / (1000^3) + (41 * (1000 - 41) / (1000^3))} = (-0.008, 0.255)$
これも0を含んでいる。

Oct. 28, 2019

47

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
- 第IV相試験: 市販後に大勢の患者が実際に服用した結果、新薬がどういう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)

Oct. 28, 2019

36

誰と誰を比べるか?

- 計画書を守った人たちだけを比べる
 - AaをA群、BbをB群
 - 2つの理由でダメ
- 実際に使われた処理で比べる
 - Aa+BbをA群、Ab+BbをB群
 - 計画書を守っていないのでダメ
- 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
 - これが正しい
 - なんで?

Oct. 28, 2019

40

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、Multiple Imputationなどの統計手法がある(伝統的な、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)
- ランダムでない欠損はいかんともしがたい

Oct. 28, 2019

44

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
 - 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
 - 通常は、新薬は従来薬より有効でなくてはならない
 - 例外: 従来薬の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 → 「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

Oct. 28, 2019

48

スクリーニング

- スクリーニング (Screening) とは？
- 「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」
 - 集団を対象に
 - すばやく実施可能な方法で
 - 暫定的に識別
 - 暫定識別なので、過剰診断は避けるべき
- (目的)一死亡率を下げる (普通では発生率は不変)
 - 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」

Oct. 28, 2019

49

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？

Oct. 28, 2019

50

スクリーニングにおけるバイアス

- 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
 - 検診の効果が過大評価される
- リードタイムバイアス:
 - 無症状での検出は症状が出る前(そのギャップをリードタイムと呼ぶ)
 - 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
- レングスバイアス:
 - 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
 - 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

Oct. 28, 2019

51

スクリーニング実施の原則 (1) (Wilson JMG and Jungner G, 1968)

- 目的とする疾患が重要な健康問題である
 - その疾患の頻度が高い (eg. がん検診)
 - 緊急な対処が必要 (eg. 新生児代謝異常検査)
- 早期に発見を行なった場合に、適切な治療法がある
 - 治療法がないと「負のラベリング効果」(例外あり)
- 陽性者の確定診断の手段、施設がある
- 目的とする疾病に潜伏期あるいは無症状期がある
- 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
 - 「適切な」は、費用や判定に要する時間も含む

Oct. 28, 2019

52

スクリーニング実施の原則 (2) (Wilson JMG and Jungner G, 1968)

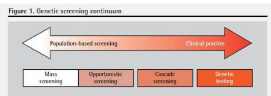
- 検査方法が集団に対して適用可能で、受け入れやすい
- 目的とする疾病の自然史がわかっている
- 患者、要観察者に対する追跡 follow-up システムが確立している
- スクリーニング事業の費用一便益が成立する
- スクリーニングの意味、内容が受診者に周知されている
 - 過度な期待をしない
 - 不要な争いを避ける

Oct. 28, 2019

53

スクリーニング実施の原則 (3)

- 現代における新しい基準 (Andermannら, 2008年)
- 遺伝的スクリーニング
 - 遺伝情報の特殊性 (玉井, 2006年): 不変性・予見性, 個人特異性, 家系内共有性 (eg. 「重い障害を持つ子どもがいる。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。似たような状態にある人はみんな受けてるじゃないか、と言い訳を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」)
 - 特有の倫理的問題: 優生思想につながりやすい。健康増進と障害者福祉の極端な場合の衝突 (DALYsの問題点と絡む)
- インフォームドチョイスにおけるプライマリケア医の重要性



Oct. 28, 2019

54

スクリーニング方法の評価 (1)

- 有効性 (感度, 特異度) と信頼性 (再現性), 迅速性, 簡便で安価であることが求められる
- 有効性の中では、一般に、陽性となる人を精密検査するための前段階として行うので、感度が高いほうがいい
- 有効性を評価するには確定診断 (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) が必要
- 感度 (sensitivity) = positive in disease
特異度 (specificity) = negative in health と覚えるといい。
- 疾病 健康
陽性 a b
陰性 c d
感度 = a/(a+c)
特異度 = d/(b+d)
陽性尤度比 = (a/(a+c))/(b/(b+d)) = 感度 / (1 - 特異度)
陰性尤度比 = (d/(b+d))/(c/(a+c)) = 特異度 / (1 - 感度)

Oct. 28, 2019

スクリーニングの評価 (2)

- 的中度は、そのスクリーニング方法の感度, 特異度のほか、対象となる集団における有病割合に依存する
- 疾病 健康
陽性 a b
陰性 c d
陽性反応の中心度 (positive predictive value) = a/(a+b)
陰性反応の中心度 (negative predictive value) = d/(c+d)
* 稀な疾病では (a+c) < (b+d) なので感度と特異度がともに高くても陽性反応の中心度が低くなることもある
- 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
 - 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

Oct. 28, 2019

56

スクリーニングの評価 (3) ~ ROC

- 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、**受検者**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, フグナー EH, 福井次夫 (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「受検者」
- 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

Oct. 28, 2019

57

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-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が結び、基準値の変化に対応する「ROC 曲線」
- 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
 - DSM-3 により何らかの精神異常があると診断された患者 40人と精神疾患がなかった60人について行った2種類の質問紙調査結果の ROC = HAD の方が優れている

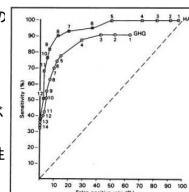


Figure 1. Relative operating characteristic curves for the hospital anxiety depression scale (HAD) and general health questionnaire (GHQ) with cut-off points indicated on the curves.

Oct. 28, 2019

58

ROC 曲線を Excel で描く例

うつの得点と臨床診断 (架空のもの)

対象者	得点	臨床診断	基準値	偽陽性率	感度
1	20	うつ	10	1	1
5	22	うつ	11	1	1
6	28	うつ	12	0.85714	1
2	13	健康	13	0.85714	1
3	19	健康	14	0.71429	1
4	21	健康	15	0.71429	1
7	11	健康	16	0.71429	1
8	25	健康	17	0.57143	1
9	16	健康	18	0.57143	1
10	19	健康	19	0.57143	1
			20	0.28571	1
			21	0.28571	0.667
			22	0.14286	0.667
			23	0.14286	0.333
			24	0.14286	0.333
			25	0.14286	0.333
			26	0	0.333
			27	0	0.333
			28	0	0.333
			29	0	0

感度 = $\frac{COUNTIF(SBS6:SBS12, "<10")}{7}$
 偽陽性率 (1 - 特異度) = $\frac{COUNTIF(SB33:SB55, ">=18")}{18}$

Oct. 28, 2019

59

母子保健分野の代表的なスクリーニング (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
 - 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
 - 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
 - 先天性代謝異常等 (ハイリスクグループ) 検出と早期治療が目的
 - 全新生児中保護者が検査を希望した者 (実際にはほぼ全数) 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
 - フェルリチン原症、楓糖尿症、ホモシチン原症、ガラクトース原症、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応
 - 平成20年度108万人達の新生児に対し、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
- 聴覚スクリーニング: 早期発見、早期治療が有効
 - 乳幼児健診での聴力検査は3歳時 (母子保健法施行規則)
 - 1990年代後半に耳音響反射 (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
 - 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。

Oct. 28, 2019

61

母子保健分野の代表的なスクリーニング (2)

- 発達障害
 - 1歳6カ月児健診で3歳児健診まで84 ~ 85%の自治体で実施
 - ただし、そのうち8割は健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
- 神経芽細胞腫
 - 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
 - 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
 - 1歳未満で見つくと予後が良く、1歳以降で見つくと予後が悪いことがスクリーニング実施の根拠
 - 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あったという知見があり、スクリーニング中止
 - 検査開始後累積罹患率が2倍になった (過剰診断)
 - スクリーニングで検出された中に自然退縮例が多い
 - 手術例の1割、化学療法例の5%に治療による合併症
 - 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が、検診事業中止により研究も中止

Oct. 28, 2019

62

成人期以降の代表的なスクリーニング (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 (前回やった)
 - 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以上かつ以下のうち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薄い
 - 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
- がん検診: がん対策基本法による
 - 国立がんセンター「がん検診ガイドライン」
 - 子宮頸がん: 細胞診のみ、相応な根拠あり
 - 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あり)

Oct. 28, 2019

63

成人期以降の代表的なスクリーニング (2)

- その他の検診
 - 脳動脈瘤・脳梗塞: 人間ドックで普通に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, γGTP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○。尿たんばくは信頼性△
 - 痛風: 人間ドックにある血清尿酸値は痛風の前期臨床症状である高尿酸血症の信頼性が◎な指標

Oct. 28, 2019

64