

# 因果関係とは？

- 要因(リスク因子)→影響(病気)の関係
  - \* その要因が原因となって、結果としての影響が表れる
- 完璧な証明は困難
  - 何らかの関連があることは相関関係で表されるが...
  - 相関では因果の向きがわからない
  - 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
  - ...等々
- もっとも強い因果関係: 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)←疫学の究極の目的は、それを明らかにすること

# 第一種の過誤・第二種の過誤

- ネイマン=ピアソン流の仮説検定
  - 母集団において「差が無い」「関連が無い」という帰無仮説
  - 「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
  - 両方を立てて統計的検定
    - [p値 < 有意水準] 帰無仮説を棄却し対立仮説を採択
    - [p値 ≥ 有意水準] 帰無仮説を棄却しない
- 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( $\alpha$ エラー:アワテモノのエラー)
- 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ことを、第二種の過誤( $\beta$ エラー:ボンヤリモノのエラー)と呼ぶ。(1 - 検出力)に等しい

# 生物学的因果関係を導く指針(不完全)

- 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・実験室的事実
- 病理学的事実
- 観察疫学的事実 = Hill の基準(時間性以外はすべて例外あり。基準というよりチェックリスト的なものだと Rothman は書いている)
  - (1)関連の強さ, (2)一貫性, (3)特異性, (4)時間性,
  - (5)用量反応関係, (6)蓋然性, (7)整合性, (8)実験証明, (9)類推

<https://www.mhlw.go.jp/shingi/2005/03/dl/s0331-9b2.pdf>

<https://www.edwardtufte.com/tufte/hill>
- 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

# 個人レベルでの因果関係は立証可能か？

- (例) 自分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て、これまで合計1年以上ソロモン諸島で過ごしたが、マラリアに罹ったことはない。
  - 靴下をはいていたから吸血予防ができて罹らなかった？
  - 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？
- 靴下をはかなかった自分は存在しないのでわからない。
- 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。
  
- 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？

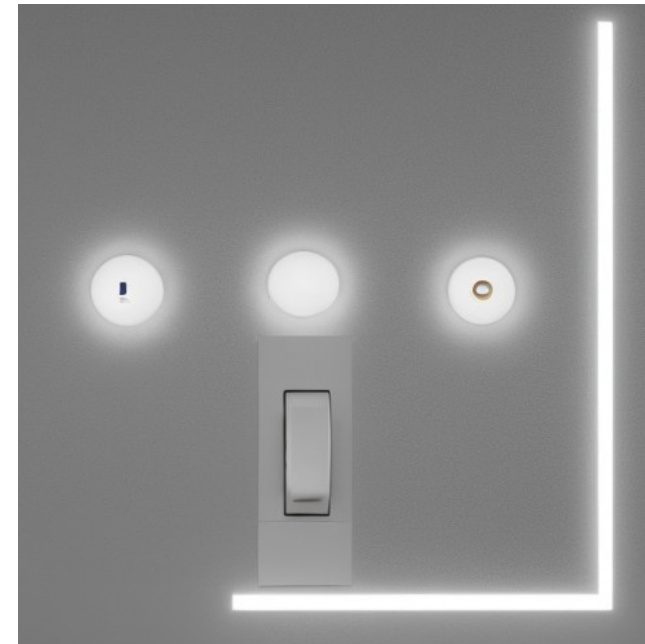
# 集団レベルでの因果推論のロジック

- 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  
\* 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可能 \*  
⇒ 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」になるように対照群を選ぶ操作をマッチングと呼ぶ。このようなコホート研究(後述)ではマッチングは重要
- (注) 症例対照研究(後述)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

# いろいろな因果関係 (1)



車に轢かれて全身打撲→交通事故死  
(注:画像は Stable Diffusion で生成)



通電していて故障していなければ、  
スイッチを入れる→照明がつく

一要因⇒一症状

# いろいろな因果関係 (2)



- 鉛の吸収過剰→腎臓, 肝臓, 中枢神経系や骨に蓄積
  - 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
  - .....

(注: 画像は Stable Diffusion で生成したイメージ)

一要因⇒多症状

# いろいろな因果関係 (3)

- 乳幼児期に B 型肝炎ウイルスまたは C 型肝炎ウイルスに感染
  - 多量飲酒
  - 喫煙
- などが相俟って、時間をかけて肝硬変→肝細胞がん

**多要因⇒一疾病**



# いろいろな因果関係 (4)

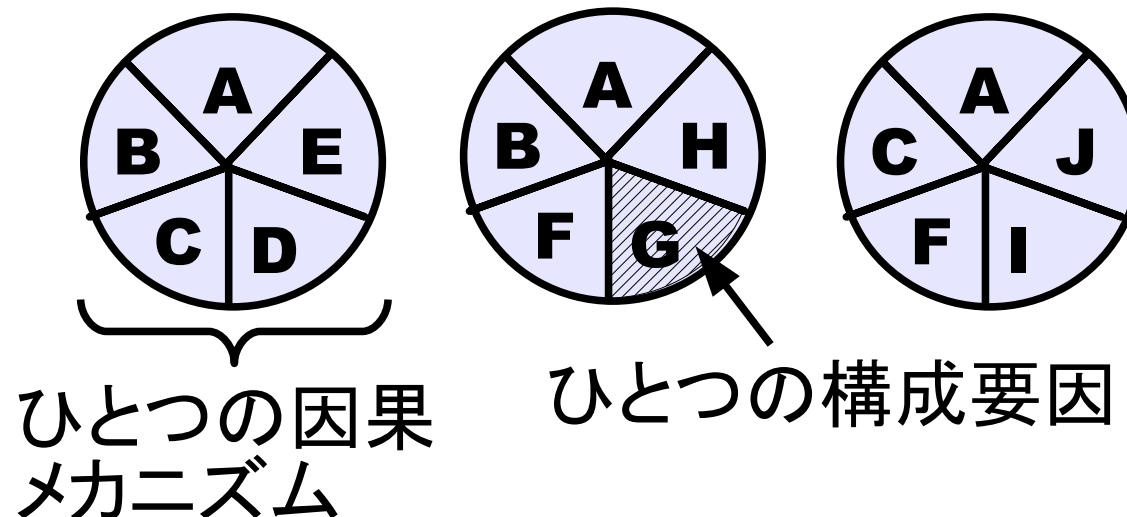
- 加齢
- 摂食過多
- 運動不足
- 遺伝因子
- 肥満
- 多量飲酒
- 喫煙
- 代謝異常

などが複雑に絡み合って、多くの非感染性慢性疾患が発症する

**多要因⇒多疾病**

# 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 )



# 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ったとする
- 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
- 1つの因果パイについてみる
  - 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)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

# 因果関係を歪めるもの

- ランダムな誤差＝測定精度が低いこと
  - 検査機器や技術の改善によって減らせる
  - サンプルサイズを大きくすることでも減らせる
- 系統的な誤差(バイアス)
  - 個々の測定値が真値から同じ向きに偏ること
  - デザインや分析により制御する
  - いろいろなバイアス
    - 選択バイアス: 観察対象が母集団の中で偏ること
    - 情報(誤分類)バイアス: 得られた情報が偏ること
      - リコールバイアス: 過去の美化・忘却・捏造
      - 公表バイアス: 意味のある結果しか公表されないため第一種の過誤が拡大される危険
    - 交絡バイアス: 交絡要因(後述)により因果関係が歪む

# いろいろなバイアスの例 (1)

- 喫煙は肺がん罹患リスクを上げるといえるか？
  - 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, 肺がんに罹ったときの致命割合が, 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(肺がん死者は, 肺がん罹患患者のうち, 喫煙者に大きく偏っている)
- 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!
- どちらも Neyman's Bias (または Prevalence-incidence bias ) と呼ばれる 選択バイアス (Hill et al., JCE 56: 293-, 2003)

# いろいろなバイアスの例 (2)

- オーケストラ指揮者は長生きか？
  - 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？
    - 確かに有名な指揮者の享年を見ると、Otmar Suitner は 87 歳、Karl Böhm は 86 歳、Herbert von Karajan は 81 歳など長生きだが……
  - 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？
  - 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
- 以上2つのバイアスを考えると、指揮が健康に良いかどうかは不明

# いろいろなバイアスの例 (3)

## • 新聞社のアンケート結果

### • 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, に良くある選択バイアス

- 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, それを書くと回答者の代表性に疑念を抱かれるので書かない確信犯)
- 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

- 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 人は, 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  
→これも選択バイアスといえる

阪神大震災で被災し、仮設住宅で暮らす人たちの中で、「復興から取り残される」と思っている人は七割を超え、震災前に住んでいた地域に「戻りたいが戻れない」と考える人も約六割に増えた。神戸市内の仮設住宅の入居者千人に対し、朝日新聞社が昨年十二月に実施した意識調査を踏まえ、うち三百人を六月に追跡調査したところ、こうした傾向が浮き彫りになった。今後の生活に不安を感じている人も増えている。一方、行政への評価は厳しさを増し、震災後の施策に対して「十分とは思わない」が八割に達した。(朝日新聞「一九九六年七月十七日」)

# いろいろなバイアスの例 (4)

- 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 = 入院率バイアス (Berkson's bias)
- 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年間観察して, 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 = 自己選択バイアス
- 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とき, 原発作業は心疾患を予防するといえるか？ 放射線ホルミシス？
  - 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ため, 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」
  - これも選択バイアスの一つ



# 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、先天異常が無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、過去について真剣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験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

# 交絡によるバイアス

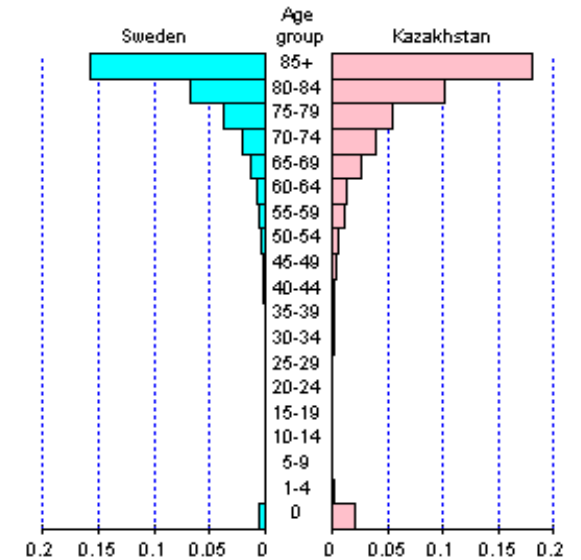
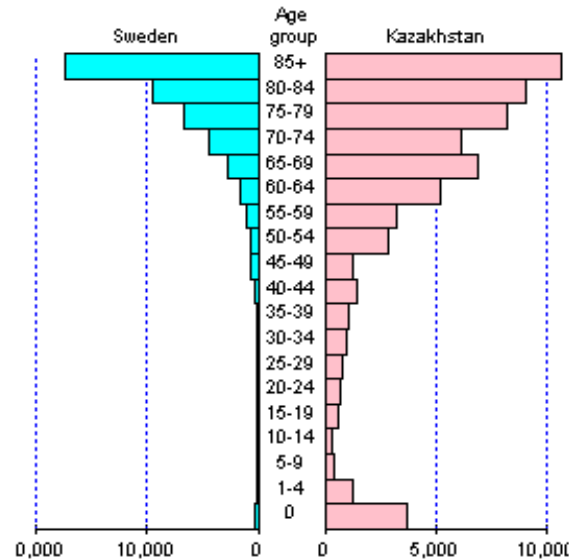
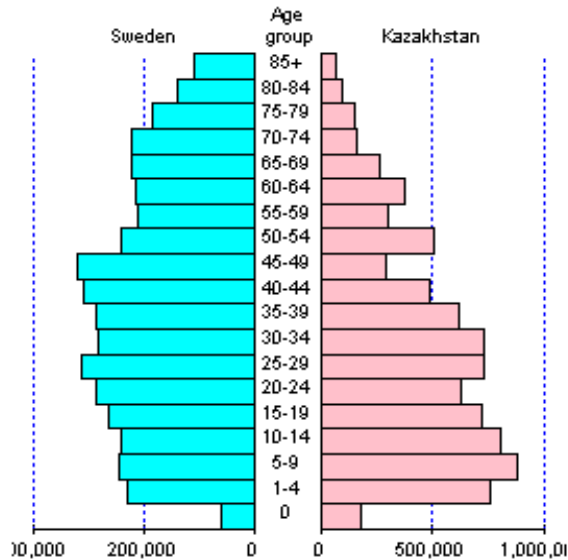
- 交絡とは、注目している因果関係を歪める第3の因子
- 交絡要因の3つの必要条件(例:肥満⇒高血圧に対する年齢)
  - 注目している要因(肥満)ではない
  - 注目している要因(肥満)と結果である健康影響(高血圧)の両方と因果的に関連
    - 「結果に影響する要因」かつ「比較したいグループの一方に偏っている要因」とも言える
  - 要因曝露の結果ではない(肥満が高齢をもたらすのではない)
    - 「原因と結果の間にある中間変数ではない」とも言える
- 制御すべき交絡を見つけるには DAG を使った裏口テストが有用(↓参考)  
[https://www.jpma.or.jp/information/evaluation/symposium/pdf/20210319/2\\_2.pdf](https://www.jpma.or.jp/information/evaluation/symposium/pdf/20210319/2_2.pdf)
- 交絡要因を見過ぐ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くなってしまふ。年齢構造が交絡要因(次のスライド)。

# スウェーデンとカザフスタンの女性の比較

Population structure of Sweden and Kazakhstan females in 1992

Number of death by age in Sweden and Kazakhstan females in 1992

Age specific death rates of Sweden and Kazakhstan females in 1992



- 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、年齢階級別死亡数、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)

# 交絡の制御

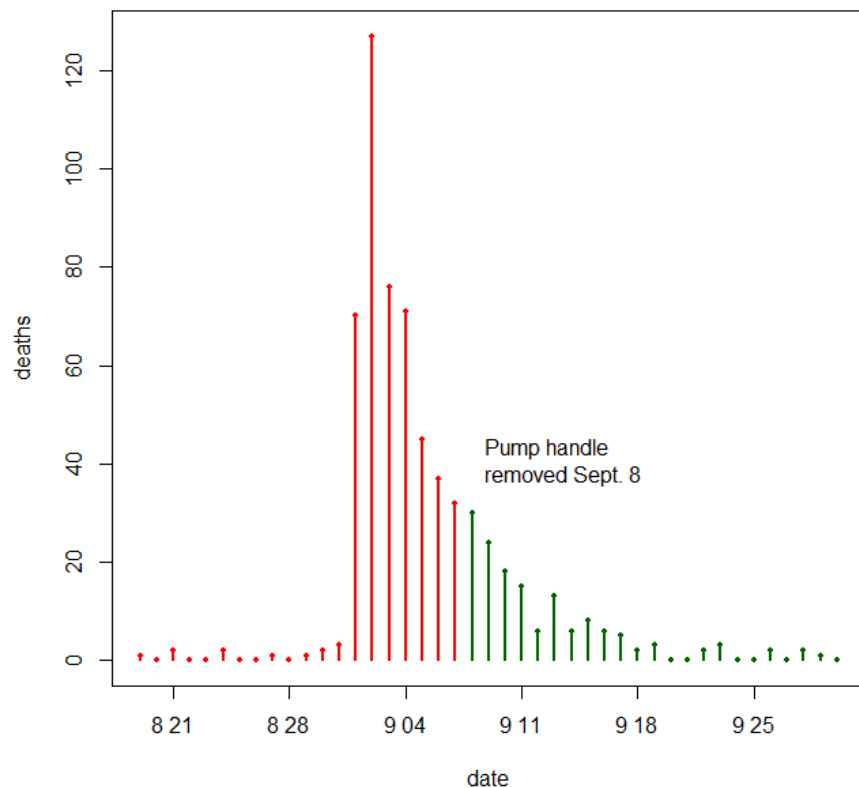
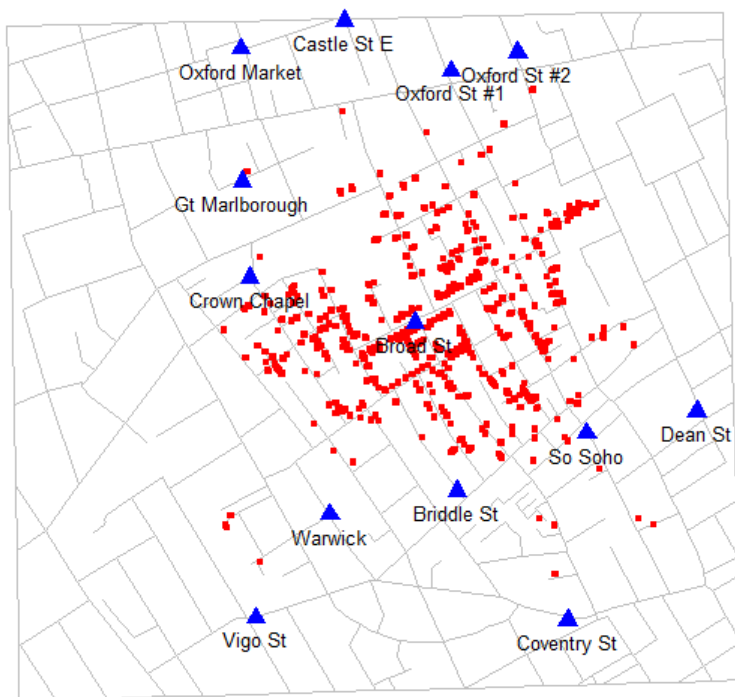
- 交絡因子によって層別し, 別々に解析
  - 年齢による交絡があるなら, 年齢階級別に死亡率を比べてみるなど
- 交絡因子の影響を制御
  - 標準化
  - プール化 (マンテルヘンツェルの要約オッズ比などあるが高度な話なので省略)
  - 限定
- 多変量解析 (ロジスティック回帰分析など)

# 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 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) と発症日別のコレラ死者数の度数分布 (下図右) を作成 (注: 図は R の HistData パッケージで example(Snow) で描画) ⇒ 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&V 社: 4093/266516 と約 5.8 倍) を計算して、S&V 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。



# 代表的な疫学研究

- 大規模コホート研究
  - Frammingham Heart Study (フラミンガム心臓研究)
  - 久山町研究 ([http://www.epi-c.jp/e001\\_1\\_0001.html](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 ([http://www.epi-c.jp/e114\\_1\\_0001.html](http://www.epi-c.jp/e114_1_0001.html))
- 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

# フラミンガム心臓研究



- 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, 子供世代コホート(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/>



# 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+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位で欧米と違い脳出血が脳梗塞の 12.4 倍とされていた→欧米の研究者から死因への疑い→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: 脳出血による死亡は脳梗塞による死亡の 1.1 倍(第 1 集団、1964 年には 100% 剖検)。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

# 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 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， $Na/K \text{ 比} \propto$  血圧，塩分摂取量  $\propto$  血圧，ヤノマミなど低塩分摂取集団では加齢に伴う血圧上昇なし，等



# 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

→人を対象とする生命科学・医学系研究に関する倫理指針

- [https://www.lifescience.mext.go.jp/bioethics/seimeikagaku\\_igaku.html](https://www.lifescience.mext.go.jp/bioethics/seimeikagaku_igaku.html)
- 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
  - 2021年からヒトゲノム・遺伝子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, プライヴァシーへの配慮が必要
  - 個人情報保護法 (<https://www.ppc.go.jp/personalinfo/>) と関連
- 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, 連結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
  - 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

# 二次資料の解析のため倫理審査不要な研究例

*Epidemiology*, 27(3): 316-322, 2016.

[<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4820668/>]

**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, concern 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 ages  $\leq 18$  years. The first round of screening included 298,577 examinees, and a second round began in April 2014. We analyzed the prefecture results from the first and second round up to December 31, 2014, in comparison with the Japanese annual 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 = 5.0; 95% confidence interval [CI] = 2.5, 9.0). The prevalence of thyroid cancer was 605 per million examinees (95% CI = 302, 1,082) and the prevalence odds ratio compared with the reference district in Fukushima Prefecture was 2.6 (95% CI = 0.99, 7.0). In the second screening round, even under the assumption that the rest of examinees were disease free, an incidence rate ratio of 12 has already been observed (95% CI = 5.1, 23).

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

Toshihide Tsuda,<sup>a</sup> Akiko Tokinobu,<sup>b</sup> Eiji Yamamoto,<sup>c</sup> and Etsuji Suzuki<sup>b</sup>

TABLE 2. Prevalence, Prevalence Odds Ratios (POR), and Incidence Rate Ratios (IRR) in Each District up to December 31, 2014

Areas and Districts (1) to (9)	Prevalence of Thyroid Cancer Cases per 10 <sup>6</sup> (95% CI)	Internal Comparison	External Comparison
		POR (95% CI)	IRR <sup>a</sup> (95% CI)
Nearest area (1) (2011 fiscal year)	359 (201, 592)	1.5 (0.63, 4.0)	30 (17, 49)
Middle area (2012 fiscal year)	402 (304, 522)	1.7 (0.81, 4.1)	33 (25, 43)
North middle district (2)	237 (123, 414)	1.0 (0.40, 2.7)	20 (10, 35)
Central middle district (3)	605 (302, 1,082)	2.6 (0.99, 7.0)	50 (25, 90)
Koriyama City district (4)	462 (299, 683)	2.0 (0.87, 4.9)	39 (25, 57)
South middle district (5)	486 (210, 957)	2.1 (0.7, 6.0)	40 (17, 80)
Least contaminated area (2013 fiscal year)	332 (236, 454)	–	28 (20, 38)
Iwaki City district (6)	451 (282, 682)	1.9 (0.84, 4.8)	38 (24, 57)
Southeastern least contaminated district (7)	236 (95, 486)	1 (reference)	20 (7.9, 41)
Western least contaminated district (8)	305 (146, 561)	1.3 (0.49, 3.6)	25 (12, 47)
Northeastern least contaminated district (9)	0 (0, 595)	0.00 (0.0, 2.6)	0.00 (0.0, 50)

<sup>a</sup>The IRRs were based on diagnosis by cytology. When based on histologically confirmed cases that were operated on, the IRRs for external comparisons using a latent duration of 4 years were 28 (95% CI = 15, 47) in the nearest area (excluding one benign case), 30 (95% CI = 22, 39) in the middle area, and 16 (95% CI = 10, 24) in the least contaminated area for which the secondary examination of cytology positive cases is incomplete.

# 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。

# 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
- 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

# 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)

# 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 で計算する場合は, `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

# 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

# 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ります、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)

# 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)

# 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)

# 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付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。

# ITT の理屈

(出典: <http://www.kbs.med.kyoto-u.ac.jp/01Sep15.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付けるとして,
  - 割付けが A, 実際に使ったのが  $A = Aa$
  - 割付けが B, 実際に使ったのが  $A = Ba$
  - 割付けが A, 実際に使ったのが  $B = Ab$
  - 割付けが 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?  
……ちょっと考えてみてください

# 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(治療効果が出てやめた人が除外される、有害作用が出てやめた人が除外される)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
  - 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？



# なぜ割付け通りに比べるか

- 割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。
- 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(実際に使われた処理で比べると, 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

# Intention-to-treat ではどうか

- 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
- 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
  - 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  
⇒ 第一種の過誤を守れる

# 割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付けられるが, 倫理など様々な理由で割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
- 最初の割付通りに解析する(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)
  - 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

# 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

# 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, 偽薬を飲んでいて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

# 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,  
 $1 - (32/1000) / (41/1000) = 1 - 32/41 = 9/41 = 0.2195\dots$   
\* 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  $1 / (9/1000) = 1000/9 = 111.11\dots$   
\* 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

- リスク比の点推定量は 0.78 で、その信頼区間は、  
 $\exp(\ln(32/41) \pm 1.96 * \sqrt{(1/32 - 1/1000 + 1/41 - 1/1000)})$   
 $= (0.495, 1.223)$   
( R の fmsb を使えば riskratio(32, 41, 1000, 1000) で出る)  
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、期待値は 0.22 かもしれないが 0 の可能性もある
- 絶対リスク減少率の信頼区間は、  
 $0.009 \pm 1.96 * \sqrt{(32 * (1000 - 32) / (1000^3) + (41 * (1000 - 41) / (1000^3))}$   
 $= (-0.008, 0.255)$   
( R の fmsb を使えば riskdifference(41, 32, 1000, 1000) で出る)  
これも 0 を跨いでいる

# 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



# 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)→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」

# 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？

# 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ため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より前（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

# スクリーニング実施の原則 (1)

## (Wilson JMG and Jungner G, 1968)

<https://www.who.int/bulletin/volumes/86/4/07-050112bp.pdf> )

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. 「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. 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. )
  1. 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. 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. )

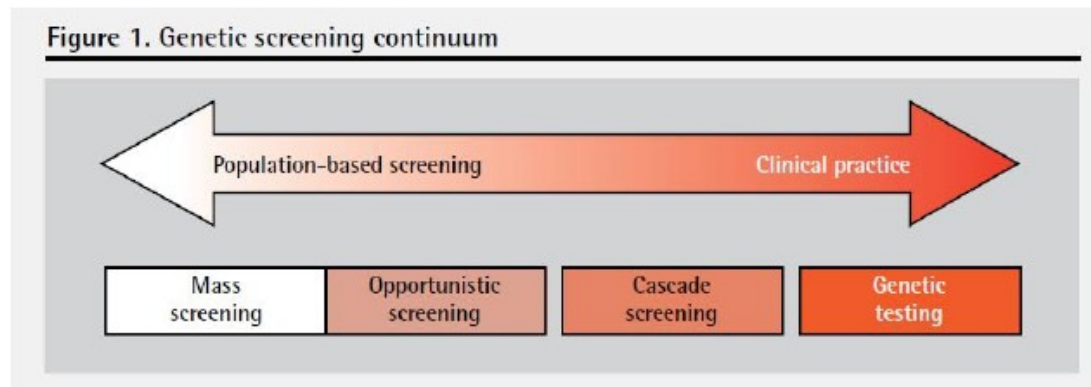
# スクリーニング実施の原則 (2)

## (Wilson JMG and Jungner G, 1968)

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. )
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. )

# 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  
<https://www.who.int/bulletin/volumes/86/4/07-050112.pdf> )
- 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, と言いつ事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



# スクリーニング方法の評価 (1)

- 有効性あるいは性能(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)によって疾病の有無がわかっている対象者について, 当該検査方法による陽性・陰性の判定が必要
  - 確定診断に使う検査方法自体の感度や特異度は? 複雑な話
- 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 =  $(a/(a+c))/(b/(b+d)) = \text{感度} / (1 - \text{特異度})$   
陰性尤度比 =  $(d/(b+d))/(c/(a+c)) = \text{特異度} / (1 - \text{感度})$

# 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

陽性反応的中度 (positive predictive value) =  $a/(a+b)$

陰性反応的中度 (negative predictive value) =  $d/(c+d)$

\* 稀な疾病では  $(a+c) \ll (b+d)$  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 = 実際的一致率と偶然的一致率の差を、1と偶然的一致率の差で割った値。完全一致のとき 1, 偶然的一致と同じとき 0, それ以下で負



# 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

# 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で結ぶと, 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
  - 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

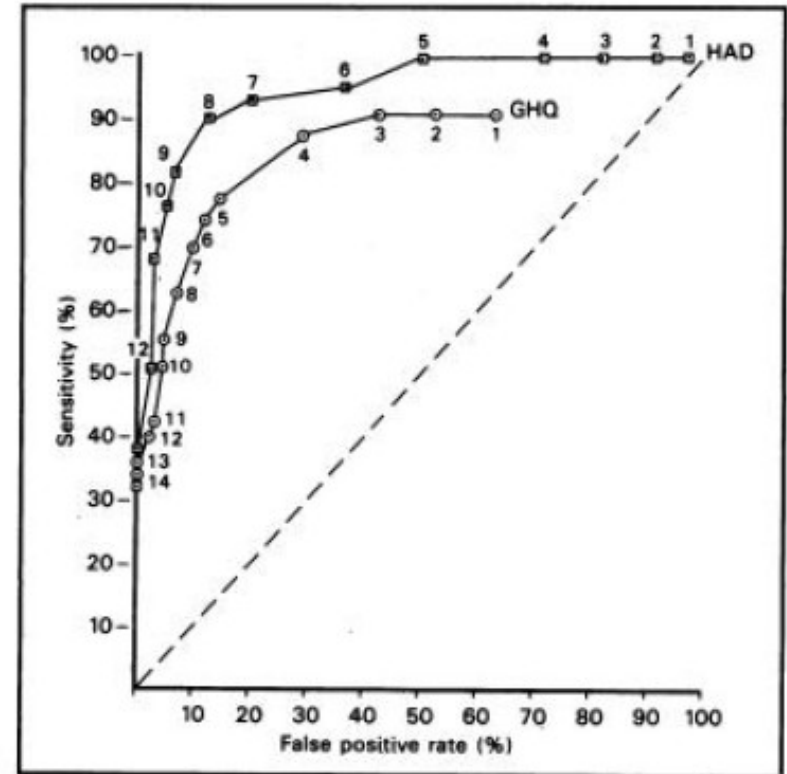


Figure 1. Relative operating coefficient curves for the hospital anxiety depression scale (HAD) and general health questionnaire (GHQ) with cut-off points indicated on the curves.

\* 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 GHQ より優れている

# ROC 曲線を Excel で描く例

うつの得点と臨床診断(架空のもの)

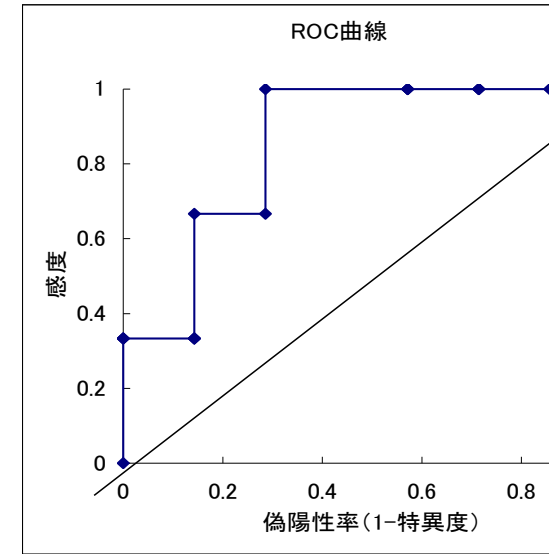
対象者	得点	臨床診断	基準値	偽陽性率	感度
1	20	うつ	10	1	1
5	22	うつ	11	1	1
6	28	うつ	12	0.85714	1
2	13	健康	13	0.85714	1
3	19	健康	14	0.71429	1
4	21	健康	15	0.71429	1
7	11	健康	16	0.71429	1
8	25	健康	17	0.57143	1
9	16	健康	18	0.57143	1
10	19	健康	19	0.57143	1
			20	0.28571	1
			21	0.28571	0.667
			22	0.14286	0.667
			23	0.14286	0.333
	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

$=1-\text{COUNTIF}(\$B\$6:\$B\$12,"<10")/7$

$=\text{COUNTIF}(\$B\$3:\$B\$5,">=18")$

基準値18のとき  
うつ 健康

陽性	3	4
陰性	0	3
感度	1	
特異度	0.429	
偽陽性率	0.571	



# RでROCを実行する方法

- データの与え方(先のデータを使うとして)

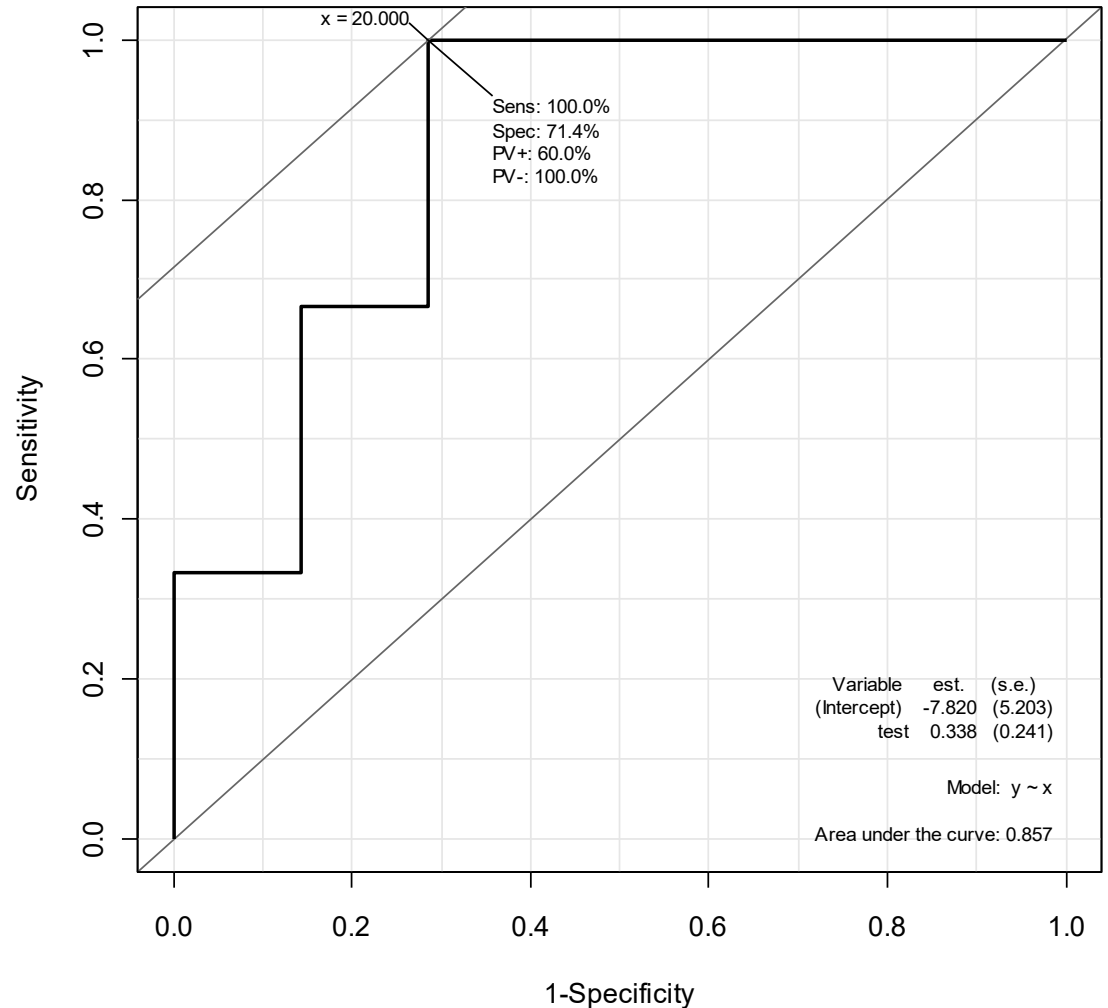
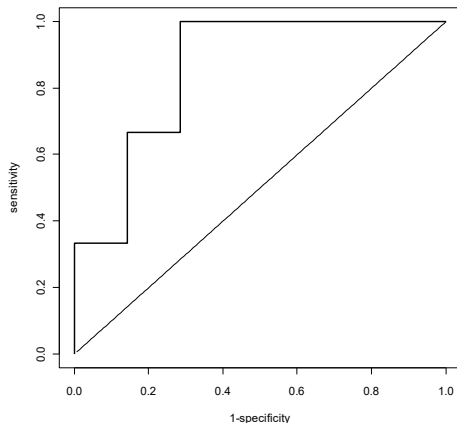
```
x <- c(20,22,28,13,19,21,11,25,16,19)
y <- c(1,1,1,0,0,0,0,0,0,0)
```

- Epiパッケージを使う場合

```
require(Epi)
ROC(x, y, plot="ROC")
```

- fmsbパッケージを使う場合

```
require(fmsb)
res <- roc(x, y)
print(res)
plot(res)
```



# 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
    - (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
  - 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
- 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
  - 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。

Gallaudet University

# 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カ月児健診でも3歳児健診でも84～85%の自治体で実施
  - 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
- 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
  - 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
    - 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

# 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
  - 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
- 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あり)

# 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