

仮説検定の統計学的基礎とその他の基本事項

- 帰無仮説と対立仮説
- 効果量・変動度
- α エラー(偽陽性)と β エラー(偽陰性)
- 有意水準と検出力(パワー)
- P値
- 多重仮説検定と事後仮説

効果量 (effect size)

- 研究者が示したい効果・関連の大きさ
- 効果量は、研究者が比較を行うために、変換する
- 研究者が決定する
 - 臨床的に意味のある値
 - 先行研究からの情報
 - 小規模なパイロット研究
- 効果量が小さいほど必要なサンプルサイズは大きくなる

効果量の指標

- 差, 比, オッズ比, 率比・・・
 - 論文では大きく見える値が載っていることもある
- 先ほどのSUMMIT研究を例に取ると,
 - MMN群: 35.5死亡/1000出生(3.55%), IFA群: 43死亡/1000出生(4.3%)
 - 絶対差 $3.55-4.3=-0.75\%$
 - 比 $3.55/4.3=0.82$
 - 相対差 $(3.55-4.3)/4.3=-0.18$ 18%の減少
- サンプルサイズの計算方法
 - 指標によって異なる
 - 結果はほぼ同じ
 - 仮説が複数ある時は一番大事なものに注目

Statistical analyses

We calculated the sample size on the basis of proportions,¹² since estimates of death rates per person-time were not available, in an intention-to-treat analysis assuming an estimated neonatal mortality ratio of 46 per 1000 livebirths because this figure would accommodate hazard ratios of early infant mortality to 12 weeks post partum. The design effect from clustered randomisation was estimated to be 1.20 for 300 clusters and anticipated loss to follow-up was 10%, with a 15% allowance for abortions and stillbirths. With a power of 80% and a two-sided significance level of 5%, enrolment and follow-up of 36 500 pregnant women and infants would be needed to detect at least a 15% change in the neonatal or neonatal mortality in the MMN group relative to the IFA group. Stopping rules drafted by the data and safety monitoring board established a minimum meaningful effect as 10%, thereby dictating a provisional sample size of 83 100 enrollees to assess effect of MMN on infant deaths.

推定の乳児死亡率:
46死亡/1000出生

少なくともMMN群が
IFA群よりも死亡率に
おいて、15%の変化を
検出する必要性

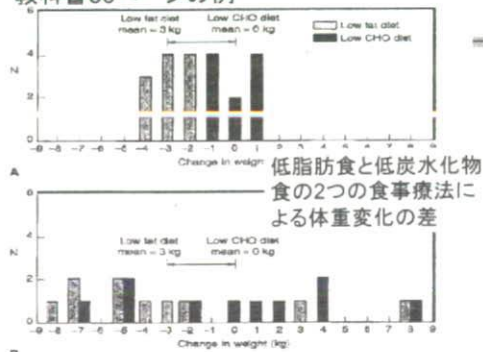
安全評価委員会は、試験中止基準は10%とした。

研究者が設定した効果量⇒2つの治療の相対差が15%

データの変動度

- 効果のばらつき具合
 - ばらつきやすい指標=サンプルがたくさん必要
- 教科書(60ページ)の例
 - 同じ効果量(平均体重差5kg)でも、データのばらつきにより、結論できる場合とできない場合がある
- できるだけ安定した指標, 安定した測定を選ぶ
 - SUMMIT研究では、「出産後90日以内の死亡」

教科書60ページの例



Statistical analyses

We calculated the sample size on the basis of proportions,¹⁷ since estimates of death rates per person-time were not available. In an intention-to-treat analysis assuming an estimated neonatal mortality ratio of 46 per 1000 livebirths because this figure would accommodate higher ratios of early infant mortality to 12 weeks post partum. The design effect from clustered randomisation was estimated to be 1.20 for 300 clusters, and anticipated loss to follow-up was 10%, with a 15% allowance for abortions and still births. With a power of 80% and a two-sided significance level of 5%, enrolment and follow up of 36 300 pregnant women and infants would be needed to detect at least a 15% change in the neonatal or postneonatal mortality in the MMN group relative to the IFA group. Stopping rules drafted by the data and safety monitoring board established a minimum meaningful effect as 10%, thereby dictating a provisional sample size of 83 100 enrollees to assess effect of MMN on infant deaths.

「10%は経過観察できない」、「15%は中絶または死産する」と予想

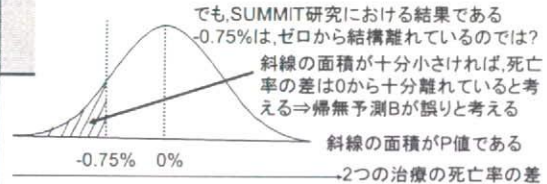
死亡という測定は安定した指標と考えられるが、ランダム化後に全員把握できるわけではない

観察される効果量の差(1):

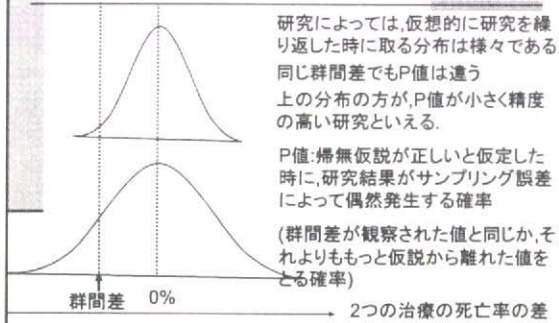
例えばSUMMIT研究の2つの治療の死亡率の差

帰無仮説A:「MMN投与とIFA投与では胎児死亡・乳児死亡は変わらない」が正しければ、

仮想的に同じ規模の研究を繰り返した時に、「2つの治療による死亡率の差」は0を中心に左右対称にばらつくはず。下記のような0を中心とする正規分布になるはず。



観察される効果量の差(2):
2つの治療の死亡率の差



結論を誤る確率

- 対象集団が母集団をうまく代表していないかもしれない
- 測定にはランダム誤差がある
- エラーをいくつまで許容するかを事前に設定する
P値が事前に定めた値(有意水準)よりも小さければ、仮説を棄却する

		母集団の状態	
		効果がない	効果がある
研究の結論	帰無仮説を採用	正しい	β エラー
	対立仮説を採用	α エラー	正しい

α エラー(偽陽性)と β エラー(偽陰性)

- α エラー(偽陽性)
 - 母集団では(本当は)、関連や効果がないのに、偶然データで効果が示してしまう確率
 - あわて者の α (効果があるととちりした)
 - α エラーをする確率が α で、統計学的有意水準と言う
 - 有意水準は、5%や1%とすることが多い
- β エラー(偽陰性)
 - 母集団では(本当は)、関連や効果があるのに、偶然データで効果が示せない確率
 - ぼんやり者の β (効果を見逃した)
 - β エラーをする確率が β で、 $(1-\beta)$ を統計学的パワー(検出力)という
 - β は、20%や10%とし検出力を80%や90%に設定することが多い

Statistical analyses

We calculated the sample size on the basis of proportions,¹⁷ since estimates of death rates per person-time were not available. In an intention-to-treat analysis assuming an estimated neonatal mortality ratio of 46 per 1000 livebirths because this figure would accommodate higher ratios of early infant mortality to 12 weeks post partum. The design effect from clustered randomisation was estimated to be 1.20 for 300 clusters, and anticipated loss to follow-up was 10%, with a 15% allowance for abortions and still births. With a power of 80% and a two-sided significance level of 5%, enrolment and follow up of 36 300 pregnant women and infants would be needed to detect at least a 15% change in the neonatal or postneonatal mortality in the MMN group relative to the IFA group. Stopping rules drafted by the data and safety monitoring board established a minimum meaningful effect as 10%, thereby dictating a provisional sample size of 83 100 enrollees to assess effect of MMN on infant deaths.

検出力:80%
有意水準:5%
両側検定を採用

MMNIによる乳児死亡に対する効果を評価するためのサンプルサイズは83,100人と見積もった

多重性の問題

- 同時に複数の仮説を検定することを、多重仮説検定という
 - 多重仮説検定を行うと偶然有意になる確率(α エラー)が高くなる
- 複数の仮説が、独立な仮説とし、有意水準を5%とすると、偶然に仮説が1つでも有意になる確率は、
仮説の数が2つならば、 $1-0.95^2=0.0975$ (9.75%)
3つならば、 $1-0.95^3=0.142625$ (14.26%)
4つならば、 $1-0.95^4=0.18549375$ (18.55%)
仮説が20個あると、1つでも偶然有意になる確率は64%まで到達する!!

多重性の対策

- ボンフェローニ法(Bonferoni method)
 - それぞれの有意水準を「 α /検定回数」にする
 - 検定が厳しくなる分サンプルはたくさん必要
 - ゲノム研究で、多数の遺伝子型と病気の関連を調べるなどに用いられている
 - でも、同じ仮説を別々の研究で検定する時には、有意水準は調整しないのに、ここまで厳しくする必要はあるの?
→ 筆者達は、臨床研究には、少し厳しすぎると考えている
- 有意水準を厳しくする必要性は、
- 偽陽性の確率が高い場合
- 事前確率が低い場合
一般に、どの有意水準を用いるかは、仮説の数よりも、検定する仮説の事前確率の方により依存する
もちろん、P値の小さい結果だけ報告するのは本末転倒。

複数の仮説を立ててはいけないのか?

- 欠点
 - 多重性の問題
 - 結論が不明瞭
 - 利点
 - 多くの研究テーマを扱える、効率が良い
 - 似通った仮説を同時に示すことで妥当性がある
- 対策
主仮説(primary hypothesis)を決めておく

SUMMIT研究においても、

Another primary outcome of the study was maternal mortality. On the basis of the above conditions and assuming an estimated maternal mortality ratio of 400 per 100,000 livebirths with an effect size of 30%, the projected sample size was 126,000. However the data and safety monitoring board indicated in March, 2004, that maternal deaths in the study had fallen to 274 per 100,000 livebirths, which was substantially lower than the ratio for which the trial was designed. This factor and the termination of the study limited conclusions about maternal mortality.

統計学的解析
の第2段落

仮説「MMN投与は、IFA投与よりも胎児死亡・乳児死亡を減らす」以外にも、仮説「MMN投与群の方が、母体死亡率が低い」も立てたけれども、デザイン通り進まず解析を断念した

Data mining (データ発掘)

- 予期しない関連
 - 偶発の産物の可能性

解釈には慎重であるべきである。報告の中心に差し替えてはいけない。
- 新しい発見
 - 新しい発見の可能性は秘めているため、将来の研究テーマになりうる。研究者が興奮するのも当然。

「新しい発見の可能性は秘めているが、その発見の妥当性は担保されていない」

まとめ (1)

- サンプルサイズの計算は、分析的研究でも記述的研究のどちらにおいても重要で、早期に見積もるべきである
- 分析的研究、実験的研究では、統計学検定を行う前程として、予測因子とアウトカムの間の関係について仮説を立てておく必要がある (純粋な記述的研究では、比較を伴わないため不要である)
- 良い仮説とは、1) サンプルング方法や測定方法が明確、2) 単純である、3) 研究開始前に立てたものである

まとめ (2)

- 帰無仮説は、予測因子とアウトカムの間には関連が「無い」とする仮説で、統計学的有意性の検定の基礎である。対立仮説とはその反対で、関連が「有る」とする仮説である。仮説検定とは、帰無仮説が否定されるかどうかを検定するものであり、否定されれば、対立仮説を採用する
- 帰無仮説には、片側のもの(関連が一方のみ)と両側のもの(関連が両方向)がある

まとめ (3)

- サンプルサイズとは、与えられた(調べられた)効果量と変動度の持つ関連を、定めた第I種過誤(α エラー、偽陽性)と第II種過誤(β エラー、偽陰性)の確率のもとで、統計学的に有意に検出するのに必要な推定対象者数のことを言う
- 研究立案段階で、複数の仮説を立てることは、研究の効率上望ましいことが多いが、研究の焦点を明確にして、サンプルサイズ計算を可能にするために、1つの主仮説を立てておく必要がある。多重仮説検定の結果や予期せぬ仮説検定の結果の解釈には、それが真である事前確率の大きさを考慮する必要がある

本日覚えるべき単語

- 帰無仮説・対立仮説
- α エラー(偽陽性)・有意水準
- β エラー(偽陰性)・検出力
- P値
- 効果量
- 変動度
- 多重検定・事後仮説・事前確率

Estimating Sample Size and Power: Applications and Examples

サンプルサイズの見積もり、その応用と例題

Designing Clinical Research 3rd Ed
Chapter 6

National Center for Child Health and Development
臨床研究センター勉強会 2008/2/5
臨床研究フェロー 前川貴伸

本日の内容

- 臨床研究の計画を立てる際に、サンプルサイズを前もって見積もることの大切さを理解する(5章の内容の復習)。
- サンプルサイズを見積もる手順を知る。
- 二つの主要な2群間比較の検定法を知る。
検定, カイ二乗検定
- 教科書6章の例題を用いて、サンプルサイズの実際の算出を経験する。

なぜ事前にサンプルサイズを見積もる必要があるのか

- 目的と予想を明確にするため
 - 先に仮説を設定することが大変重要。
 - なにを知りたいか?それは意味のあることか?
- Power(検出率)を試算するため
 - サンプルサイズが小さいと結論が出せない
 - (試験によるが)80%以上の検出率がほしい
- 実施可能性を考慮するため
 - サンプルサイズが大きいと実施が難しくなる
 - 研究が動き出したら、途中の変更は難しい

折笠秀樹,『臨床研究デザイン—医学研究における統計入門』

サンプルサイズの見積もりの手順 2群間の差を検証する場合

- ①「差がない」という仮説を設定(帰無仮説)
 - すべてはここからはじまる
- ②アウトカムの種類(連続変数,2値変数か)から必要な検定法を判定する。
 - アウトカムの種類により,2群の差の示し方(検定法)が異なる
- ③効果量(Effect Size)を見積もる
 - アウトカムの差がどれくらい予想されるか
- ④効果量を標準化する
 - 非常に大切な考え方
- ⑤検出率(Power)の水準を設定する
 - これは自分で設定する

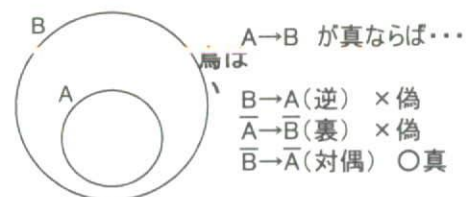
①帰無仮説の設定

2群の比較:2群に差があることを示したい。

基本的アプローチ

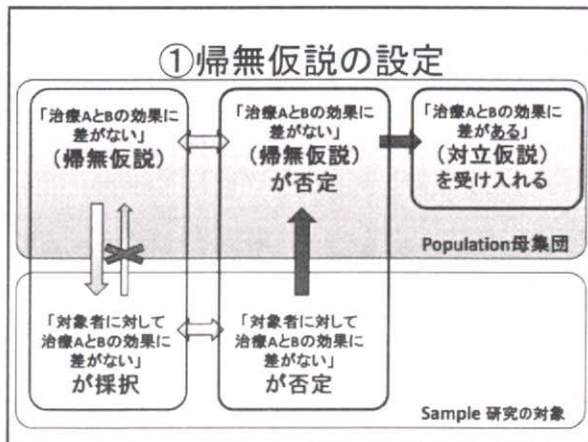
- ある2群の対象者について結果に差があったことを示すのは意外と大変。
 - ①「どれくらいの差」なのか
 - ②ある対象では真(差があった)と言っても、他の対象でもそうなのか
- ↓
- 「帰無仮説」(差がないという仮定)を用いる

白鳥は全て白い?



A:世界で白鳥は全て白い Ā:世界では、白くない白鳥がいる
B:日本で白鳥は全て白い B̄:日本では、白くない白鳥がいる
(Not All swan is white)

A:母集団で帰無仮説が採択 Ā:母集団で帰無仮説が棄却
B:サンプルで帰無仮説が採択 B̄:サンプルで帰無仮説が棄却



② アウトカムの種類を知る

予測因子 Predictor Variable	アウトカム変数 Outcome Variable	
	2値変数 Dichotomous	連続変数 Continuous
2値変数 Dichotomous	X ² 検定	t検定
連続変数 Continuous	t検定	相関係数

Table 6.1 Simple Statistical Tests for Use in Estimating Sample Size (p66)

実際には、アウトカムのデータ数、分布、分散関連性の有無により、シンプルなStudent t検定が適応できないこともあります。

③ 効果量 (effect size)を見積もる

効果量
アウトカムを評価のできる形に変換したもの

- ◆ 研究で捉えたいと考えている因果関係における要因と結果の「関連の強さ」
- 2群間の治療効果の差、比など
例) 平均血圧降下、死亡率の差、オッズ比...
- 研究者が見積もる。先行研究、小規模なパイロット研究などから推定するが、これがなかなか大変な作業。
- 効果量が小さいほど必要なサンプルサイズは大きくなる

④ 効果量の分布を標準化する

◆ 2群間のアウトカムの差を標準化する

効果量 / 標準誤差
Effect Size / Standard Error (SE)

- ◆ T検定 2群間の差 / 2群間の差の標準誤差
- ◆ カイ二乗検定 割合の差 / 2群間の割合の標準誤差

標準誤差で除すことにより、サンプルが大きい場合、この値が標準正規分布に従う性質を利用する。

⑤ α エラー、 β エラーの水準を設定

- ◆ α エラー(擬陽性)
 - 本当は関連や効果がないのに、効果が見えてしまう確率
 - あわて者の α (効果があると早とちりした)
 - α エラーをする確率が α で、統計学的有意水準と言う
 - 有意水準は、5%や1%とすることが多い
- ◆ β エラー(擬陰性)
 - 本当は関連や効果があるのに、効果が見えない確率
 - ぼんやり者の β (効果を見逃した)
 - β エラーをする確率: β
 - 見逃さない確立 = 検出力(Power): $(1-\beta)$
 - β は20%や10%と設定することが多い(Powerが0.8か0.9)

統計学的検定法

平均値, 標準偏差, 標準誤差

例) 1年1組女子30人の体重を測定した。

結果: 平均体重20.5kg 標準偏差4.07kg

- 日本人小学1年女子の平均体重を知りたいとき, 30人から得られた“20.5kg”という数値をどれくらい信頼してよいのか?

横断的標準体重曲線(2000年度版)
6歳6か月 女子平均体重 21.3kg 標準偏差 3.4kg (男子平均体重 21.8kg)

t分布のパーセント点(自由度29)
 $p=0.4(0.854)$ $p=0.3(1.055)$ $p=0.2(1.311)$ $p=0.1(1.699)$ $p=0.05(2.045)$

標準偏差と標準誤差

- ◆標準偏差(Standard deviation; SD)
 - データのバラツキを示す指標
 - 集団全体(たとえば日本人の身長など)の分布のバラツキを表現する要約統計量。
- ◆標準誤差(Standard Error; SE)
 - 推定値のバラツキを示す指標
 - 推定値の標準偏差
 - この例題の場合, 30人の平均体重のバラツキの指標が平均値の標準誤差(SE)である。

$$\text{平均標準誤差 } SE = SD / \sqrt{n}$$

平均値, 標準偏差, 標準誤差

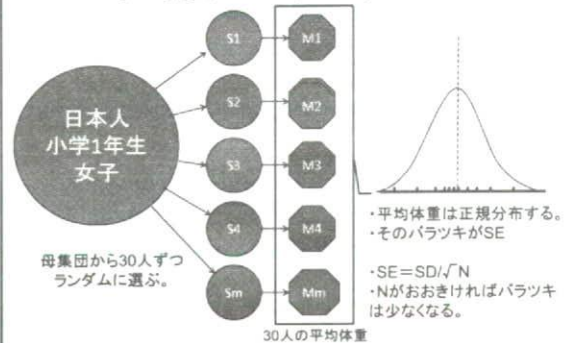
例) 1年1組女子30人の体重を測定した。

結果: 平均体重20.5kg 標準偏差4.07kg

- 平均標準誤差
 $SE = SD / \sqrt{n} = 4.07 / \sqrt{30} = 0.74$
これが平均値のバラツキ
- 95%信頼区間を知る
自由度(30-1)=29のt分布では, $p=0.05$ に対応する $t=2.045$
 $20.5 \pm t(SE) = 20.5 \pm 1.51 = 18.99 \sim 22.01$
95%の確立で母集団の平均体重はこの信頼区間に入る。

横断的標準体重曲線(2000年度版)
6歳6か月 女子平均体重 21.3kg 標準偏差 3.4kg (男子平均体重 21.8kg)

平均値のバラツキとは?



アウトカム変数が連続変数の場合

- ◆呼吸機能検査における一秒量(例題b-1)
- ◆体重, 身長(先ほどの例)
- ◆血液検査データ
- ◆血圧などのバイタルサイン

など多岐にわたる

t検定

- アウトカム変数が連続変数の場合に, 2群間のアウトカムの平均値が有意に異なるかどうかを検定する場合に用いられる。

【前提】

- アウトカムが(ほぼ)正規分布する
✓でも, 正規分布にしたがわないほとんどの分布についても使用できる

【例外】

- サンプルが小さすぎる($n < 30-40$)
- 極端な外れ値

t検定

- アウトカム変数が連続変数の場合に、2群間のアウトカムの平均値が有意に異なるかどうかを検定する場合に用いられる。

【前提】

- アウトカムが(ほぼ)正規分布する
 ✓でも、正規分布にしがわかないほとんどの分布についても使用できる

【例外】

- サンプルが小さすぎる($n < 30-40$)
- 極端な外れ値

Sample A

サンプル数 N_1

平均値 M_1

分散 V_1

標準偏差 SD_1

平均標準誤差 SE_1

Sample B

サンプル数 N_2

平均値 M_2

分散 V_2

標準偏差 SD_2

平均標準誤差 SE_2

この2群が同じ母集団からとられたSampleだと信じられるか？

そこで、「2群の平均値の差」に注目する。

• 同じ母集団からとってきたSampleならば、「2群の平均値の差」は0の周りに分布するはずである。

2群A,Bの「2群の平均値の差」が0より大きく
 かけ離れていれば、AとBは同じ母集団から
 とってきたSampleだと信じられない

自由度 $(N_1 + N_2 - 2)$ の t 分布

「2群の平均値の差」の分布

「2群の平均値の差」/ 差の標準誤差
 は自由度 $(N_1 + N_2 - 2)$ の t 分布に従う。

計算式

差の標準誤差: $SE(\text{Diff})$

$$\sqrt{(V/N_1 + V/N_2)}$$

分散: V

$$\{(N_1 - 1)V_1 + (N_2 - 1)V_2\} / (N_1 + N_2 - 2)$$

例題6-1 気管支喘息発作治療薬の効果比較 サルブタモールとイプラトロピウム

- 予測因子: サルブタモールとイプラトロピウム
 > 2値変数
- アウトカム変数: 治療1週間後のFEV₁ (1秒量)
 > 連続変数
- 2値変数 × 連続変数
 > t 検定により効果を検証する。

例題6-1

t検定

- 帰無仮説の設定
- 効果量 (E: effect size) を見積もる: 評価項目平均値の群間差がどの程度か?
- 評価項目のバラツキvariabilityを見積もる: SD
- 標準化した効果量 (E/S) を計算
- α と β を設定

例題6-1

t検定

Sample Size per Group for Comparing Two Means

One-sided $\alpha =$	0.05		0.025		0.05		0.10	
Two-sided $\alpha =$	0.01		0.05		0.10		0.20	
E/S	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10
0.10	3,365	2,579	2,338	2,020	2,180	1,971	2,166	1,938
0.15	1,998	1,525	1,242	1,117	1,007	906	993	920
0.20	1,111	793	636	561	512	452	507	458
0.25	772	479	378	317	288	253	307	275
0.30	566	350	270	228	207	181	227	199
0.35	427	265	198	166	151	131	166	143
0.40	321	197	148	125	114	98	128	108
0.45	241	146	109	91	83	71	95	81
0.50	181	109	80	67	61	52	69	58
0.55	137	80	60	50	46	39	51	43
0.60	104	60	45	38	35	29	38	31
0.65	79	45	33	28	26	21	28	23
0.70	60	35	26	22	20	16	21	17
0.75	48	28	21	18	16	13	17	14
0.80	38	21	16	13	12	10	13	11

* E/S is the standardized effect size, computed as f (average effect size) divided by s (SD of the outcome variable). To determine the standardized effect size, first compute the standardized effect size and then divide the specified value of α by f for the required sample size in each group.

②アウトカム変数が2値の場合

- ◆生存／死亡
- ◆疾病の発症／非発症
- ◆再発／無再発

検定方法:カイ二乗検定

- 2値変数のアウトカムをもつ2群間の差を検定する方法。
- 2値変数は、数値化しないと比較ができないので、例えば、
 - ◆(病気ある/なし)→ 罹患率の差
 - ◆(生/死) → 死亡率の差
 といった具合に、その効果(Effect size)は「割合の差」として表現される。

例)タモキシフェンと乳がん再発

	再発		
	あり	なし	
TAM+	140	797	937
TAM-	172	764	936
			1873

Abe O. Breast Cancer 1994; 1:1-9

TAM投与群 $P_+ = 140/937 = 0.149$ (14.9%)
 TAM非投与群 $P_- = 172/936 = 0.184$ (18.4%)

効果量 $(P_+) - (P_-) = -0.035$ (-3.5%)
 これは意味のある差なのか？

例)タモキシフェンと乳がん再発

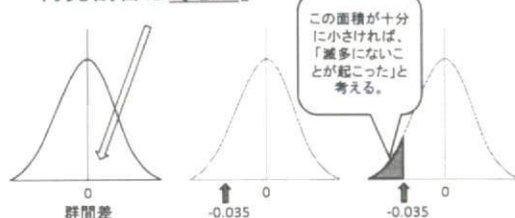
【検定手順の概念】

- 帰無仮説
「TAM投与と乳がん再発は無関係」
- ↓
- 帰無予測
「サンプルにおいて、TAM投与群と非投与群で再発割合は等しい」= 群間差はない

例)タモキシフェンと乳がん再発

• 帰無予測

「サンプルにおいて、TAM投与群と非投与群で再発割合は等しい」



例)タモキシフェンと乳がん再発

• 帰無予測

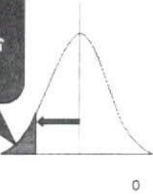
「サンプルにおいて、TAM投与群と非投与群で再発割合は等しい」が棄却

- ↓
- 帰無仮説
「TAM投与と乳がん再発は無関係」を棄却
- ↓
- 対立仮説
「TAM投与すると乳がんの再発割合が減少する」を採択

P値

「群間差が観察された値と同じか、それよりもっと仮説から離れた値をとる確率」

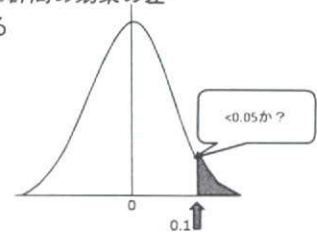
全体に対する
この面積の割合
(%)がP値



正規分布では
±1SDの間に
86%が分布し、
±1.96SDの間
に95%が分布。

例題6-2 カイ二乗検定

- ① 帰無仮説を設定 2群間で皮膚発生率は同等
- ② アウトカムを確認 皮膚癌ある/なし=2値変数
- ③ 効果量を確認 2群間の効果の差
- ④ α値、β値を設定する



例題6-2 カイ二乗検定

	皮膚癌		
	あり	なし	
喫煙あり	X_1	$(N_1)-(X_1)$	N_1
喫煙なし	X_2	$(N_2)-(X_2)$	N_2
	$(N_1)+(N_2)$		

$P_1 = 0.30$ $P_2 = 0.20$

$P_1 - P_2 = 0.10$

$\alpha = 0.05$ $\beta = 0.20$

$$z = \frac{(\hat{P}_1 - \hat{P}_2) - 0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

$p = 0.5$ で標準
誤差最大
 $N \uparrow \rightarrow SE \downarrow$

例題6-2 カイ二乗検定

表 6-11 2群の割合を比較する場合に必要なサンプルサイズ

PとPの差 存在いか?	PとPの差					
	0.05	0.10	0.15	0.20	0.25	0.30
0.05	26	176	72	42	22	17
0.10	17	136	66	36	21	16
0.15	16	125	62	32	19	15
0.20	14	108	57	28	18	14
0.25	13	98	54	26	17	13
0.30	12	91	51	24	16	12
0.35	11	84	48	22	15	11
0.40	10	78	45	20	14	10
0.45	9	73	42	18	13	9
0.50	8	68	40	16	12	8

2群間の割合を比較する場合に必要なサンプルサイズ (表6B)

相関係数

- 予測因子とアウトカム因子がともに連続変数である場合
- 散布図が書ける
- 相関係数(r)を用いて評価。

$A \leftrightarrow B$

- 相関を示すのに必要なサンプルサイズは？

$A \leftrightarrow B$

- 帰無仮説
「AとBは相関関係がない」 = 「 $r = 0$ 」
- 帰無仮説を棄却するのに必要なサンプルサイズ

予測因子:連続変数
アウトカム因子:連続変数

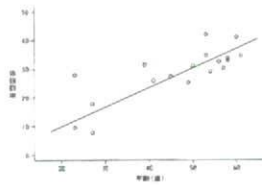


図 11.1 日本人の成人女性における年齢と関連する体脂肪率
(Kawachi, 1996)

二つの方法

- 想定する相関係数を効果量(r)を決め、相関係数が0でないことを示す(帰無仮説を棄却する)に必要なサンプルサイズを算出。
- 説明変数を2値変数に置換
 > 50歳前後で体脂肪率分布に差があるか
 • t検定の原理に従いサンプルサイズを算出

脱落(Dropouts)

- 脱落を見越してサンプルサイズを算出する必要がある。

- 脱落が20%と見積もり



- サンプルサイズは125%必要
($125 \times 0.8 = 100$)

カテゴリー変数

- ◆ 年齢(65歳未満、65歳以上)
- ◆ 低出生時体重(<2500g/2500g \leq)
- ◆ ペインスケール(0,1,2,3,...,10)

- 連続変数はカテゴリー変数に置換可能
- カテゴリーが多い(≥ 6)場合は、連続変数とみなして扱うことができる。

生存分析

- 生存分析
- 標準薬Aと新薬Bで延命効果に差はあるか?
- アウトカム「延命効果」は連続変数(6ヶ月、1年、3年など)のように見えるが、実際には各時点での生存割合をみている

	6か月	1年	3年
A	60/100	50/100	45/100
B	80/100	50/100	45/100

→ χ^2 分析の原理に従って、サンプルサイズを算出する。

マッチング

- 2群間比較をする場合、背景情報の似ているサンプリングを行う。
- これにより、サンプル内で生じうるバラツキが少なくてできる可能性がある。
 - 20歳と60歳の女性のある薬剤の効果の比較
 - 20歳代の女性サンプルにおける薬剤の効果の比較

サンプルサイズを少なくする方法

- アウトカムとして、連続変数を用いる
- より精度の高い(バラツキが少ない)アウトカムを指標とする
- より頻度の高いアウトカムを用いる

まとめ

- サンプルサイズの見積もりの手順に従う。
- 効果の指標を標準化する手順が重要。
- 標準化効果量が帰無仮説の棄却域にあることで対立仮説を採用する。
- 脱落を考慮するほか、検出力を上げる工夫も必要。
- 記述研究、生存分析、クラスターサンプル、多変量解析についてはあらためて講義があります。

薬物動態学の基本的概念

薬剤部 小村 誠

生理学的モデル

- 投与部位
 - 脈管内: 動脈、静脈
 - 脈管外: 経口、舌下、口腔、筋肉内、皮下、皮内、経皮、肺、直腸・・・など
- 薬物の処理
 - 肝臓: 薬物を代謝して無毒化
 - 腎臓: 未変化体、又は代謝物を尿中に排泄

生理学的モデル

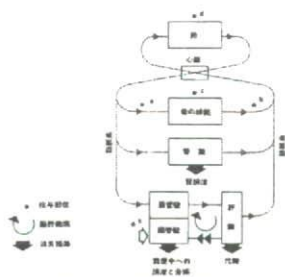


図 2-1 全身平均薬物動態学 (physiological modeling) と投与部位 (site of administration) x 代謝 (metabolism) の関係 (relationship between site of administration and metabolism) x 排泄 (excretion) の関係 (relationship between site of administration and excretion)

薬物の科学的特性

- 立体異性体 (鏡像異性体) を有する薬物
 - ワルファリン: S体とR体のラセミ体
 - S体の方がR体よりも活性が高く、半減期が短い
 - S体の代謝酵素はCYP2C9、R体はCYP1A2及び3A4と、異性体により代謝酵素が異なる
- 放射性医薬品
 - 放射活性の尿中回収率 = 薬物の尿中排泄率?
 - > 薬物動態学でいう尿中排泄率は「未変化体」の回収率をいう

吸収 (absorption)

- 投与部位から生体内における測定部位 (全身循環系、一般的には血中又は血漿) まで薬物が未変化のまま運ばれる過程として定義
 - 消化管での未吸収
 - 消化管での分解 (代謝): CYP3A4
 - 肝臓での代謝
- 初回通過効果 (first-pass effect)
 - 消失部位における最初の薬物通過時に起こる消失
 - その割合は薬物によって異なる
 - 例: ニトログリセリンの first-pass effect は 90% 以上 → 舌下投与

バイオアベイラビリティ

- 血管外投与された薬物量と全身循環血中に到達した薬物量を関係づける定数 (F)
 - 全身循環血中に到達した薬物量

$$= D (\text{投与量}) \times F$$

$$= D \times (F_a \times F_f)$$

$$F_a: \text{吸収率}$$

$$F_f: \text{初回通過効果回避率}$$
 - * F_f は、 F_g (消化管での初回通過効果回避率) $\times F_h$ (肝での初回通過効果回避率) で示すこともできる

薬物の吸収過程

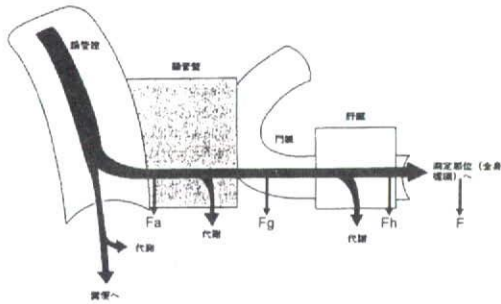
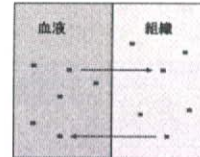


図 2-2 薬物の消化管吸収

分布 (distribution)

- 血液か血漿である測定部位(全身循環系)からの薬物の可逆的な移行過程(例えば筋肉と血液との行き来)と定義



消失 (elimination)

- 測定部位(全身循環系)からの非可逆的な消失であり、代謝と排泄の2つの過程がある
 - 排泄(excretion)は、未変化体の非可逆的な消失→腎臓、胆汁
 - 代謝(metabolism)は、他の物質への変換
 - ・ 第1相: 酸化反応、加水分解反応
 - ・ 第2相: 抱合反応
 - 腸肝循環(胆汁中に排泄→小腸から再吸収)

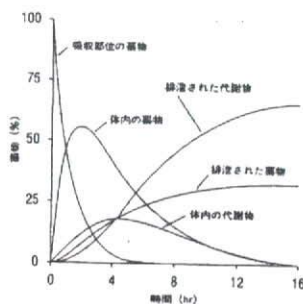
薬物動態のイメージ

- 薬物濃度の時間的推移を速度論的に取り扱う場合に、生体を単純化してモデル化したもの
 - 投与部位、体内、排泄などと明確に区別される



- * 投与量 = 吸収部位の薬物量 + 体内薬物量 + 排泄された薬物量 + 体内代謝物量 + 消失した代謝物量
- * 体内薬物量の変化速度 = 吸収速度 - 消失速度(代謝速度 + 排泄速度)

薬物動態のイメージ



吸収が100%であったときの薬物と代謝物の時間的推移

1-コンパートメントモデル

- 薬物の分布過程において、体内循環系(血中)と分布組織を1つのコンパートメントとするモデルのこと
- 血中濃度と組織中濃度が直ちに平衡になるものと考えられる
 - 多くの薬物の場合、1-コンパートメントモデルで考えられる
 - 血中濃度と組織中濃度が平衡に到達するのに時間がかかる場合、2-コンパートメントモデルという(例:ジゴキシン)

分布相と消失相

- 分布相とは、静脈内投与したときの投与部位（血中）から体内（臓器）への相を示す
- 消失相とは、血中濃度の減少が体内からの消失のみによるときの相を示す

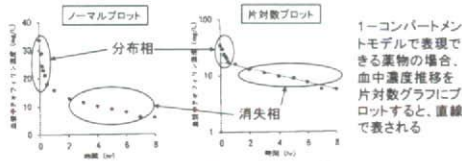


図 3-1 アナール、100 mg 静脈内投与後の血中濃度 (体重 70 kg) の血中濃度 (Miyama, P. A., J. Clin. Pharmacol. Ther., 14, 555-559, 1979 より転載)

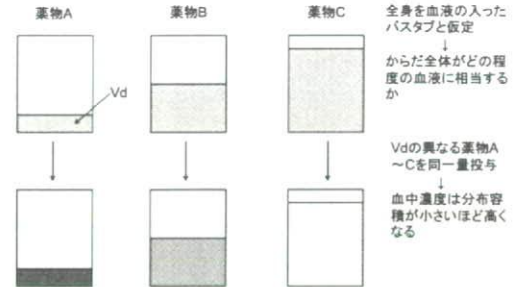
分布容積

- 血中濃度 (C_p) と体内に残存 (分布) している薬物量を関連づける比例定数 ($V_d [L, L/kg]$)
体内薬物量 (A_b) = $V_d \times C_p$
- 60kg男性の平均体液量は、血液量: 3L、細胞外液量: 12~16L、全体液量: 36~42L



図 3-7 様々な薬物の分布容積の値

分布容積

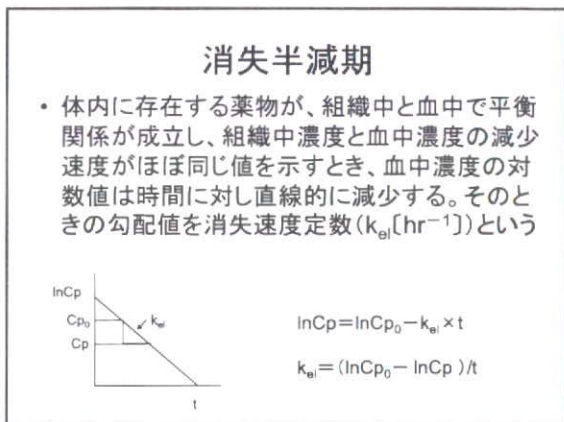
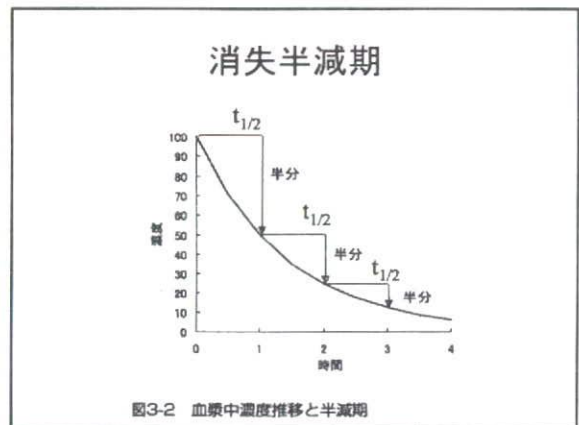
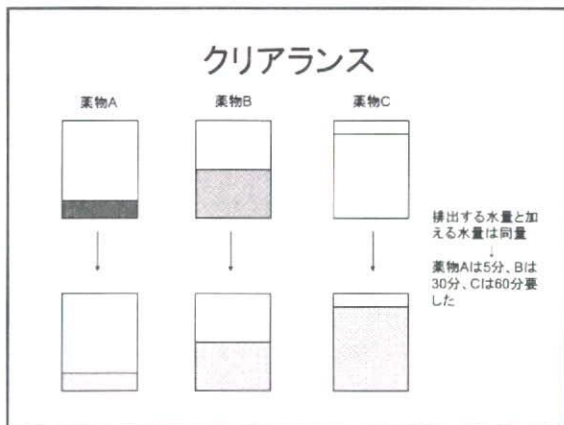


クリアランス

- クリアランスとは、薬物の消失速度と血中濃度を関連づける比例定数 ($CL [l/hr, ml/min]$)
薬物消失速度 = $CL \times C_p$
- 全身からの消失を考える場合のクリアランスを全身クリアランス (CL_{tot}) という
- CL は血中濃度当たりの消失速度を示し、同一血中濃度で比較した場合、 CL が大きいほど体内からの消失速度は大きい
- クレアチニンクリアランスも同じ考え方

クリアランス

- クレアチニンクリアランス (CL_{Cr})
- $CL_{Cr} = \frac{\text{尿中クレアチニン濃度 (Ucr)} \times \text{尿量 (U)}}{\text{血清クレアチニン濃度 (Scr)}}$
- $Ucr \times U$ はクレアチニンの消失速度を表す
- $CL_{Cr} = \text{消失速度} / Scr \rightarrow \text{消失速度} = CL_{Cr} \times Scr$



消失半減期

- $k_{el} = (\ln C_{p_0} - \ln C_p) / t$ の式より、一定時間内において血中濃度の減少率は濃度に関係なく一定であることが分かる
- 血中濃度が半減する時間(半減期: $t_{1/2}$)は、一定値を示す
 $\rightarrow t_{1/2} = 0.693 / k_{el}$

消失半減期

- Vd: 薬物に汚染された体液の架空の容積
- CLtot: 全身からの消失能(清浄能)
 \rightarrow 単位時間当たりの血中濃度(の対数値)の低下度 (k_{el}) は、汚染された体液量に対するその体液の清浄能の比であるとも考えられる
 $\rightarrow k_{el} = CL_{tot} / V_d$

まとめ

- ① 体内薬物量 (A_b) = $V_d \times C_p$
- ② 薬物消失速度 = $CL \times C_p$
- ③ $t_{1/2} = 0.693 / k_{el}$
- ④ $k_{el} = CL_{tot} / V_d$
- ⑤ ③と④から、 $t_{1/2} = 0.693 \times V_d / CL_{tot}$

観察的研究をデザインする

コホート研究
2008年2月18日
臨床研究フェロー 寺町

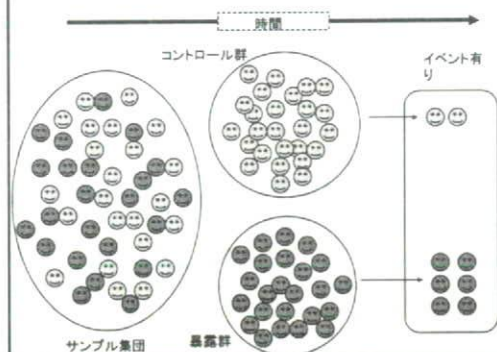
研究デザインの分類

- 観察研究 (observational study)
 - 症例集積研究 (case series study)
 - 横断研究 (cross-sectional study)
 - 縦断研究 (longitudinal study)
 - ケースコントロール研究 (case-control study)
 - コホート研究 (cohort study)
- 介入研究 (intervention study)
 - 比較対照試験 (controlled trial)
 - 対照なしの研究 (uncontrolled trial)

今日の内容

- コホート研究
 - 前向きコホート
 - 後ろ向きコホート
 - コホート内ケースコントロール研究
 - 多重コホート研究
- コホート研究を企画する

コホート研究



コホート研究でわかること

日科技連臨床試験セミナー 2006. 8 久保田

Incidence risk: (新たに発生したケースの数)/(スタート時点の集団の人数)

	イベント		合計	
	あり	なし		
あり	a	b	a+b	リスク=a/a+b
暴露				
なし	c	d	c+d	リスク=c/c+d
			a/a+b	
リスク比 (risk-ratio) =			c/c+d	

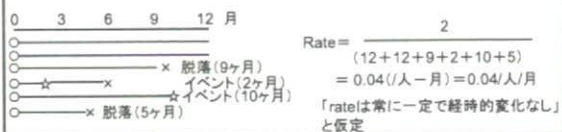
コホート研究でわかること

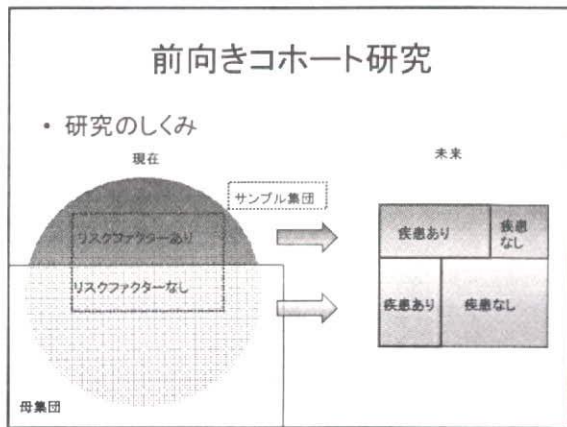
日科技連臨床試験セミナー 2006. 8 久保田

Incidence rate: 新たに発生したケースの数/総人一月(人一年)

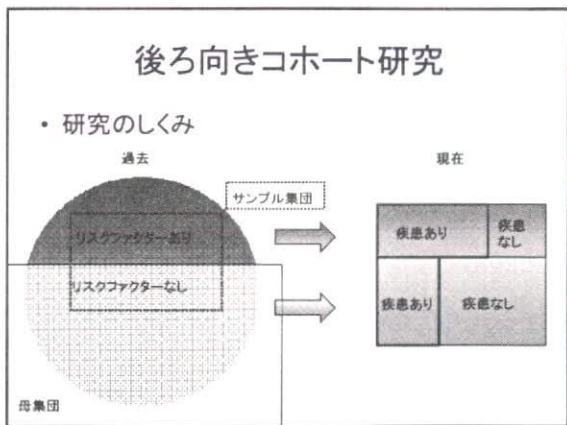
総人一月(人一年): 集団の全メンバーのtime at riskの総和

Time at risk: 期間の最初から終わりまで(非ケース)
期間の最初から脱落まで(非ケース)
期間の最初からイベント発生時点まで(ケース)
「イベントが発生したら、もはやat riskではない」





- ### 前向きコホート研究
- 利点
 - 予測因子 (predictor variable) がアウトカム発生前に測定されているため両者間の因果関係についてより確実な推論を行うことができる。
 - 重要な因子に関する測定を完全かつ正確に行うことができる。
 - 欠点
 - 稀な疾患の研究には効率が悪い。
 - コストがかかりすぎる。



- ### 後ろ向きコホート研究
- 利点
 - 前向きコホートと同様、事前に予測因子の測定が行われるため、アウトカムが発生したことを知ってから予測因子を測定する場合に生じるバイアスを回避できる。
 - 経費・時間が少なくすむ。
 - 欠点
 - 測定が過去に完了しているため、サンプリング法や予測因子の内容や質をコントロールすることができない。
 - 既存のデータに重要な情報が含まれていないことがある。

後ろ向きコホートの例として・・

Improved renal survival in Japanese children with IgA nephropathy

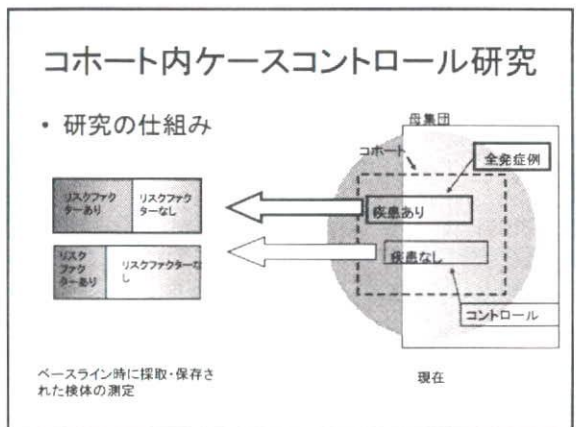
Study by: Koichi Nakase, Takashi Ueda, Shiro Ueda, Shiro Nakase, Masahiro Ueda, Kazuhiko Ueda, Kazuhiko Ueda, Kazuhiko Ueda, Kazuhiko Ueda

IgA腎症の1970年代から現在まで、500例の長期経過

IgA腎症の真のエンドポイント(腎不全)を評価

1990年代から治療法の変化あり、前期と後期での予後の違い

Group	N	CI	P	Q
1970-1989	128	95	0.0001	19.8
1990-1999	143	91	0.0001	19.8



コホート内ケースコントロール研究

- 利点
 - ベースライン時の保存検体が利用できる。
 - 測定数を限定できる。
 - 生き残りバイアス (survivor bias) を回避できる。
- 欠点
 - 他のコホート研究と同様に、得られた関連が交絡による可能性がある。
 - 進行の遅い疾患ではベースライン時に測定された予測因子が前臨床期の病態の影響をうけている可能性がある。

コホート内ケースコントロール研究

INTERNATIONAL JOURNAL OF TUBERCULAR LUNG DISEASE

ORIGINAL ARTICLE

Asthma as a Risk Factor for Invasive Pneumococcal Disease

Thomas R. Jaeger, M.D., M.P.H., Frank Hansell, M.D., M.P.H., and Michael M. C. Nahata, M.D., M.P.H., Patrick G. Arbuprod, M.D., Katherine A. Roeding, M.D., M.P.H., William Schaffner, M.D., Alfred S. Craig, M.D., and Marck Griffin, M.D., M.P.H.

2005;35(2):2082-90

コホート内ケースコントロール研究

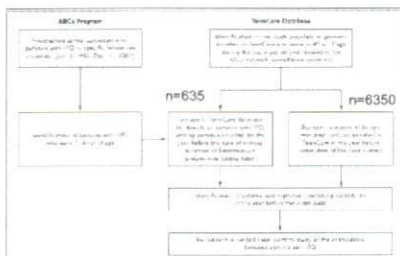


Figure 1. Diagram of the Nested Case-Control Study. From January 1993 through August 1999, the Centers for Disease Control and Prevention (CDC) Active Bacterial Core surveillance (ABC) research network and the Tennessee Cohort in Tennessee (1990-1999) identified and compared cases and controls with invasive pneumococcal disease (IPD) among the ABC program and the Tennessee Cohort program, and IPD among pneumococcal disease.

コホート内ケースコントロール研究

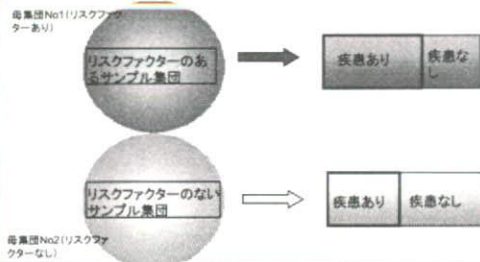
Table 1. Association between the Presence of Asthma and the Risk of Invasive Pneumococcal Disease.

Variable	Case Subjects (N=671)	Controls (N=6350)	Adjusted Odds Ratio for Invasive Pneumococcal Disease (95% CI)*
any asthma	11,633 (17.3)	15,635 (19.1)	2.4 (1.8-3.1)
High risk asthma	9,633 (14.2)	16,436 (20.9)	2.5 (1.9-3.3)
Low risk asthma	1,999 (2.9)	1,919 (2.9)	1.7 (0.96-3.0)
Controlling confounding factors a high risk of invasive pneumococcal disease†			
46-year	12,147 (17.7)	16,054 (20.4)	2.4 (1.7-3.4)
36-year	11,288 (16.7)	16,058 (20.4)	2.3 (1.7-3.1)
6-year‡	28,122 (41.3)	15,629 (19.7)	2.3 (1.4-4.0)
5-27§	1,432 (21.1)	16,430 (20.9)	4.0 (2.5-6.0)¶
18-49¶	27,791 (41.3)	16,412 (20.8)	2.4 (1.8-3.1)

* Conditional logistic regression was used to control for multiple confounding factors, and the reference of asthma was in the nested pneumococcal disease case-control study and disease risk factors without confounding factors. Data are given as odds ratios (OR) and 95% confidence intervals (CI) for each disease risk factor. † Data are for high risk asthma. ‡ Data are for low risk asthma. § Data are for asthma. ¶ Data are for asthma.

(前向き)二重コホート研究

- 研究のしくみ



多重コホート研究

- 利点
 - 稀な要因に対する暴露、職業あるいは環境由来の有害要因への暴露の影響を評価できる唯一の方法。
 - 外部コントロールとして国勢調査や疾病登録のデータを利用でき、コスト面や一般集団ベースの研究となる。
- 欠点
 - 他のコホート研究と同様

コホート研究を企画する

- どのような対象者をコホートに選ぶか
 - 追跡しやすい、研究目的にあった集団
- 研究結果の質
 - 予測因子とアウトカムの測定の定度 (precision) と真度 (accuracy) に左右される。
 - 交絡因子の測定
- 対象者のフォローアップ
- 対象者の脱落

まとめ

- コホート研究とは
 - ある疾患の発生率や自然史の記述
 - アウトカムと予測因子の関連性
- 前向き、後ろ向き研究
- コホート内ケースコントロール研究
- 多重コホート研究
- 因果関係
 - 交絡因子の測定
- フォローアップ率

飲酒と死因

