

Cohort study

Cohort study に incidence rate ratio を好んで用いる理由

Incidence rate (IR)は疾患の発生を検討する上で重要であることは先に述べました。Incidence rate ratio (IRR), incidence rate difference (IRD) は exposure と疾患発生との関係を見る上で欠くことはできません。一方 cumulative incidence あるいは risk は IR が非常に小さい値になってしまうような場合あるいは追跡期間が短いような場合に用いるべきです。追跡期間が十分であれば cumulative incidence ratio は有意差なしの方向に近づく傾向にあり、rate が低くても IRR について検討するべきです。更に competing causes of death が存在する場合 cumulative incidence ratio は過小評価してしまうため用いるべきではありません。

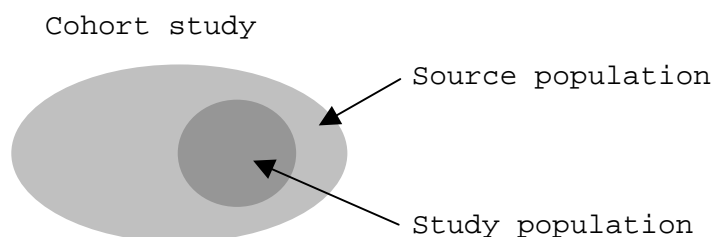
どの集団と対象に研究を行なうかー研究デザインにおいて最も重要な問題

しばしばどの集団を対象に臨床研究を行なったか不明確な場合があります。しかし、対象集団を明確にすることは臨床研究をデザインする上で最も重要なことなのです。しばしば用語が混同していますので以下用語の解説から始めます。

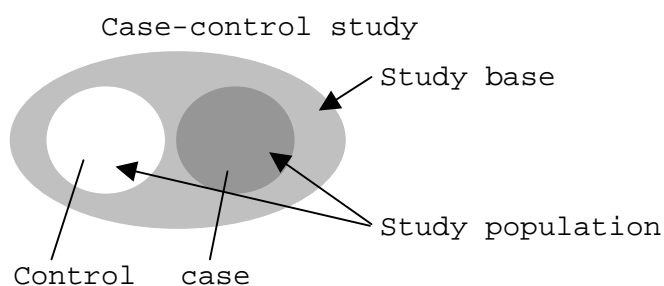
Target population: 例えば 20 代から 60 代のナースを対象に調査お行った場合、研究者は追跡調査しやすいという点でナースを選んだと思いますが、結論を全女性にまで拡張することを考えています。すなわち、結果をどこまでに適応できるか、そのターゲットが target population ということになります。閉経前の女性を対象に行なった経口避妊薬の冠動脈発生に関する調査結果から閉経後も含めた全年代層の女性を target population とすることはできません (lack of generalizability or external validity)。Effect modification が存在すると適応できないのです。

Study population: Cohort study において person time に組み込まれる人々のことを指します。すなわちある臨床研究において 1000 人の応募者中から診察の結果基礎疾患のない健康人 500 人を選んできて諸々の調査をしたとすると 500 人は study population であり、1000 人は source population ということになります。また

study population = source population である場合も時にありますが、通常の cohort study においては study population は source population の一部です。



Case control study においては case + control = study population ということになります。Case と control は study base から抽出されます。



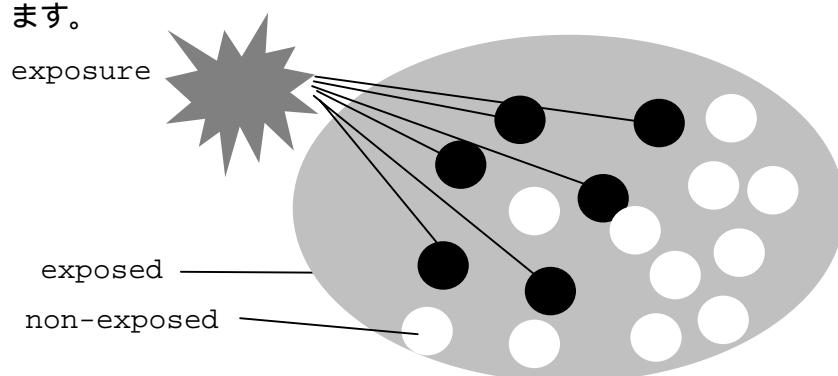
Source population: study population を抽出した源の人々を指します。Study population が source population より random に抽出された場合、study population の結果をもって source population を推論できます。例えば港区でランダムに電話連絡して1000人を抽出し港区の一般的傾向を推論する場合です。しかし、電話でコンタクトをとれる人は常に家に居るとか携帯電話を持っている人々で、必ずしも港区住民の平均にあたるかどうかは判りません。連絡する時間帯によって職種が違つかもしれません。よって一般的には study population を source population から完全に random に抽出することにこだわらないことが多いようです。このような状況を conditional occurrence relation と呼びます。

Base population: 所謂 person time になる人々を指します。ですから cohort study においては base population = study population であり、case control study においては study base の中から person time に貢献する人々を指します。

Study base: cohort study においては study base = study population = base population で、person time を構成する人々を指します。よって cohort study においてはこんなに多くの名前と呼ぶ必要はなく、かえって混乱の元のような気さえします。一方 case control study において、case と control は共通の study base から抽出します。抽出の際病気の有無のみで抽出され、exposure と独立してはなりません。

Cohort study は exposure based である。

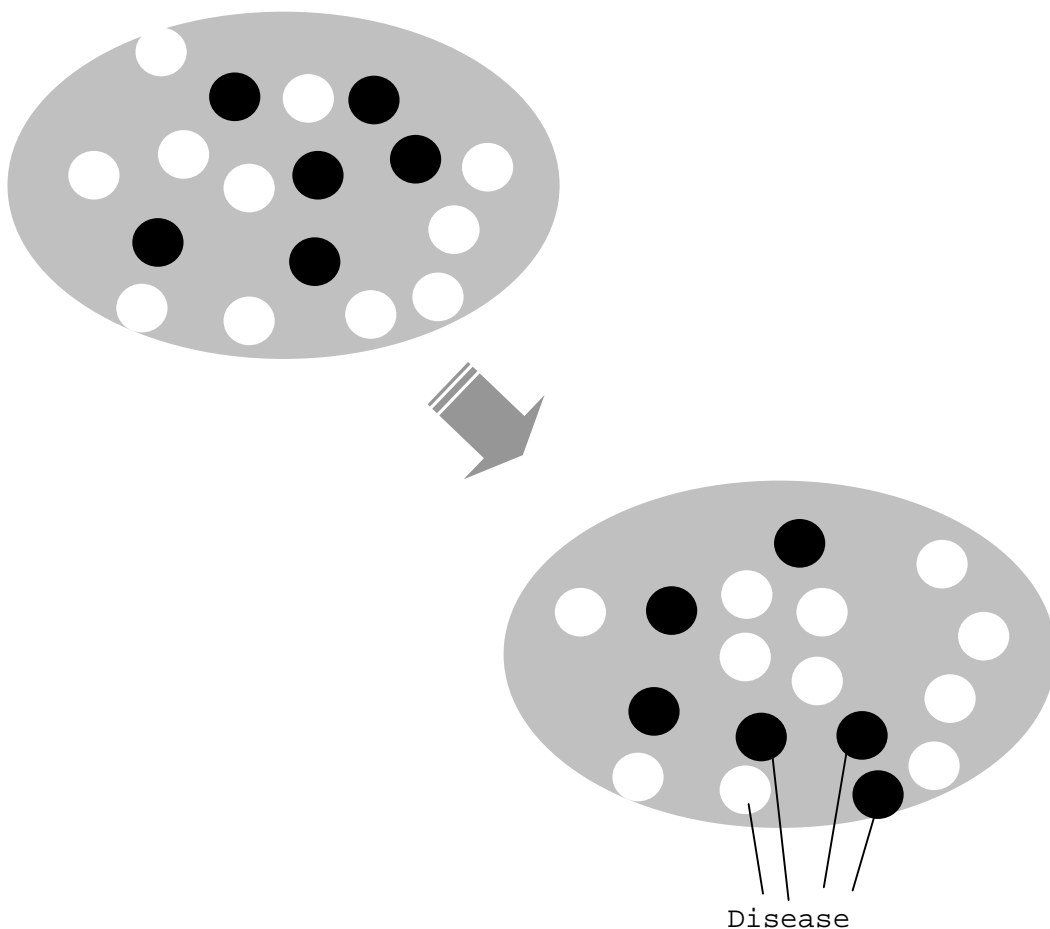
研究対象となる人は exposure の有無によって抽出されます。これに対して Case control study では結果を軸に研究するものです。先に述べたとおり結果が起こっているかないかで検討するのは retrospective か prospective の相違で、cohort study と case-control study の相違はこれとは区別されます。その exposure 自体は希な場合もあれば一般的な場合もあるでしょう。また一生継続するものもあれば可逆性のものもあります。また必ずしも放射線被爆のような暴露というのに相応しい要素から女性であるといった暴露という表現が適切でない場合まで含まれます。



Cohort study は longitudinal である

Cohort study では exposure から outcome 発生まで一定の期間を必要とします。Outcome が発生してから調査するものを retrospective, outcome が発生する以前より調査を開始するものを prospective と呼びます。しかしながら exposure 以前に outcome が発生してしまっている場合は除外されなくてはなりません。臨床研究の方法論をよくみると、大概この点についても記載があるはずです。Prospective study において exposure が起こる前から開始してもかまいません。Retrospective study の場合 recall bias を生じ易い欠点があります。最近アメリカでは保健会社

が患者データを一括してコンピュータ管理していたり、国や大病院においては記録がきちんと成されている為、研究対象によっては recall bias は非常に小さくやりやすくなりました。しかし我々は1週間前何を食べたか正確に覚えていないものです。ましてや数年前のことに關しては記憶が曖昧です。一方 Prospective study は研究内容によっては非常に時間がかかり追跡不能例が増えると研究の有効性を失う恐れさえあります。



Cohort study の型

Closed (fixed) と open (dynamic) の2種類に分けられます。

Closed cohort においては最初に study population を決定しますが、open cohort においては10年間の調査期間途中で加わることも可能であり、途中で追跡不能になってもかまいません。たとえば東海村に臨界事故のあった際生活していた人々を追跡調査する場合 closed cohort であり、それ以降移住してきた人々は含まれません。また途中追跡不能になった場合は調査対象から除外しなくてはなりません。臨床研究論

文の方法にしばしばその旨が記載されています。1万人応募して5000人が適応であり、研究参加に同意し4000人で開始したところ最後まで経過を追えたのは3000人でした、といった具合です。一方ある病院Aに1990年から2000年までの間にカルテのある人々を対象に研究する場合は open cohort になります。例えばAさんは調査の開始から終了までA病院にかかっていたが、Bさんは1995年から最後まで、Cさんは1993年から1995年までかかったとします。open cohort では3人とも対象となります。そしてこのような場合 person time として10年間追跡調査したAさんは2年間しか追跡されなかったCさんよりより大きく分母の数値に貢献することになります。また10年間の追跡調査で喫煙者が追跡開始より2年で禁煙してしまった場合でも person time に組み込むことはできますが、cumulative incidence では対象からはずさなくてはなりません。

Closed study では incidence rate と cumulative incidence を測定できますが、前者では各々の観察期間が判っていかなくてはなりませんし、後者では追跡不能者が少なく、competing risk がなく、観察期間中の exposure risk が同じであり、対象は皆同じ期間観察されなくてはなりません。例えば東海村放射能漏れ事故で多くの死亡が不幸にしてでてしまった場合、cumulative incidence で測定しようとすると白血病の発生率は真の値より過小評価されてしまいます。また他の場所へ引っ越す人々も多いかもしれません。長期追跡調査では老化も問題もからんできます。ですから closed study においても incidence rate で算出する方が有利なのです。Cumulative incidence を使うとすれば短期で結果発生の頻度が高い場合、例えば食中毒や急性感染症の流行時などの臨床研究に向いています。Study base の人数は validity には関係ありませんが、人数が多い程統計学的パワーを上げることができます。また同じ人数であれば exposed vs. unexposed の人数も均等である方がより大きなパワーを得ることができます。またパワーを上げるためには、genuine variation、つまり exposure の variance によって層化できることも大切です。たとえば喫煙の程度、あるいは過去何年間の喫煙などによって分けるといった意味です。Open cohort study では研究の性格上 incidence rate のみを測定します。

Cohort study における効果測定

	Exposure	No-exposure
Outcome	X_1	X_0
No-outcome	$N_1 - X_1$	$N_0 - X_0$
total	N_1	N_0

Cumulative incidence difference = $X_1 / N_1 - X_0 / N_0$

Cumulative incidence ratio = $(X_1 / N_1) / (X_0 / N_0)$

Odds ratio = $(X_1 / N_1 - X_1) / (X_0 / N_0 - X_0)$

疾患が希な場合 cumulative incidence ratio と Odds ratio は近似しますが、そうでない場合は Odds ratio の方が 1 (有意差なし) より離れます。多くの臨床研究において Odds ratio が好んで用いられますが、cohort と case control の間でリンクする点と logistic regression などでも OR が用いられるためだと思われます。例えばある因子が男児の出生割合を下げるとします。仮に 0.65 から 0.5 になったとすると risk ratio は $0.5 / 0.65 = 0.77$ となります。一方女児は $0.5 / 0.35 = 1.43$ となります。しかし 1.43 は 0.77 の逆数ではありません。一方 OR でみると 0.54 と 1.86 になり逆数の関係にあります。このような感覚を好むのであれば OR を用いるべきでしょう。

Closed cohort study において経過観察が長期化すれば途中死亡や経過追跡不能が増え、観察が困難になってきます。そのような場合 person-time に基く rate を考慮しなくてはなりません。

Incidence rate difference = $X_1 / T_1 - X_0 / T_0$

Incidence rate ratio = $(X_1 / T_1) / (X_0 / T_0)$

ここで T は person-time であらわされます。