

## 第3章「疾病発生と因果的効果を測る」

### 1 疾病発生を測る

測定は疫学でも中心課題。疫学が疾病（今日では広く健康関連事象を含めて解釈せねばならない）発生の研究であると定義する立場で考えれば、疫学の基礎となる観察は疾病発生の測定である。そこで本章では、いくつかの疾病発生尺度と、それらに由来する曝露と疾病の因果関係を示す可能性がある基礎的尺度を扱う。

#### 1.1 リスクと発生割合

risk

リスク。危険とも訳すが、本来的には悪い意味はない。下式の通り、疾病のリスクとは、ある期間に疾病を患う（のべ）人数を、その期間に追跡した（のべ）人数で割った値として定義される。

$$\text{Risk} = \frac{A}{N}$$

累積罹患率（cumulative incidence rate）ともいう。リスクを疾病頻度の尺度として使う第一の利点は、理解しやすいことだが、科学的尺度として使うには概念を明確にする必要がある。大事なのは追跡期間によってリスクは変わること。<sup>a</sup>

<sup>a</sup> 例)「60歳女性が心疾患で死亡するリスクが2%である」は何を意味するのか？ 次の24時間で典型的な60歳女性の2%が死亡するというのは明らかに正しくない。1年間でもまだ高すぎる。しかし残りの人生全部のうちに致死的な心疾患を患うリスクは2%より高いだろう。2%という期間が正しくなるような、適当な追跡期間があるはず。

ある期間における全てのリスクは、期間開始時点ではゼロであり、時間が経つにつれて理論的上限である100%に近づく。Fig.3-1は20年間のリスクの上がり方の2つのパターンを示す。Aは感染症に罹った人が免疫をもつ場合のように、初期にリスクが上がるけれどもその後頭打ちになるパターン、Bは胎児期のDESへの曝露によって若い女性が膵腺ガンを発症する（2章で論じた）場合のように、何らかの原因に曝露して、導入期間の最後に近いところまで観察するとみられる、着実にリスクが上がり続けるパターンである。

疾病発生を評価する尺度としてのリスクの欠点は、適切な時間間隔に渡ってリスクを測ることが技術的に不可能という点にある。ある期間観察すると、対象者の中には、必ず、問題の疾病以外の原因で死亡する人がでてしまう。例えば、1万人の既婚女性を30年追跡してDV発生を測りたいとしても、30年生きられるのは1万人全部ではない。中には極端なDVで死亡する女性もいるだろうが、心疾患やガンや感染症、交通事故などで死亡する方が遥かに多そうである。5年間追跡した間はDV被害に遭わずに死んだ女性がいたとしよう。もし彼女が生きていたら、その後の25年間にDVの犠牲にならなかったと言い切ることはできないだろう。この女性をリスクの分母に入

れてしまうと 30 年間生きた女性集団における DV リスクを過小評価することになってしまう。5 年で死ななくても、DV 被害に遭わないまま 30 年間を生き抜かず死んだ女性を分母に入れると、DV 発生リスクを過小評価することになる（分子 A は 30 年間フルに生きていたとしての DV の犠牲者をカウントすべきだから）。全死因による死亡のリスクを調べたい場合は、測っていない原因で誰かが死亡する可能性はないことになるが、それでも、原因の研究が終わって追跡調査している間に、測っていた原因を 1 つも経験せずに死んでしまう人がいる可能性はある。そこで競合リスクが問題となる。

incidence proportion

発生割合。ある集団の平均リスクをいうときに用いる語。risk と同じ意味で用いることもしばしばある。累積罹患率（cumulative incidence rate）ともほぼ同義。

competing risks

競合リスク。疫学調査の対象者が、観察している帰結（何らかの健康関連事象）とは別の帰結によって対象でなくなる場合に、後者の帰結のことをいう。例えば、喫煙と肺がんの関係の研究において、対象者が冠状動脈性心疾患で死亡してしまっ観察対象でなくなることがあるが、この場合は、冠状動脈性心疾患が競合リスクである。

loss to follow-up

追跡不能例。lost to follow-up ともいう。いかなる理由であれ研究の追跡対象者となってから研究が終了するまで継続して参加できなかった、あるいは参加しなかった例。

## 【コラム】発病率と致命率

### attack rate

発病率。流行期間内の一定期間における、観察された集団の感染の累積罹患率。この「率」は経験的に臨床例の確認や血清疫学的手法によって決定される。期間が不確かだったり恣意的に決められるので、率として記述されるべきではないとされる。

### secondary attack rate

二次発病率。初発症例に曝露された後の潜伏期間中の接触者から生じたある感染者数を分子とし、曝露されたすべての接触者の人数を分母とする割合。伝染性の指標であり、対策方法を評価する際にも有用。関連した概念として、基礎再生産率 (basic reproductive rate: 1人の患者から平均して何人の新しい患者が生み出されるかを示す値) がある。

### primary cases

初発症例。病原体が入ってきて最初に感染し発病した患者。

### case fatality rate

致命率。ある疾病に罹患した人のうち、その疾病で死亡した人の割合(％で表す)。疾病の重篤度を示す。ただし慢性疾患では有病期間が長いので、観察期間の設定が重要。いくつかの仮定をおけば、致命率 = 死亡率 / 罹患率といえる。

## 1.2 罹患率

### incidence rate

罹患率。発生率とも訳すことがある。最大のポイントは、この値が割合でなくて速度であるということである。分子の単位が件数(のべ人数)であるのに対して、分母の単位が人・時(通常は人年)なので、罹患率の単位は時間当たりになる。意味としては、瞬時における病気へのかかりやすさ。つまり瞬時における疾病罹患の危険度(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 / 罹患率となる。

### time at risk of disease

疾病リスクの時間。発生率の分母のためにどうやって時間を数え上げるかについての統一的な概念は単純: 分母に入る時間は、その間に研究されている疾病あるいはイベントが起こる可能性がある、追跡されている人々が経験する時間に対応している。そこで、発生率の分母に数え上げられる時間をこう呼ぶことがしばしばある。

mortality rate

死亡率。イベント発生が死亡によって測定される発生率のこと。

person-time

人・時。人年 (person-year), 人月 (person-month), 人日 (person-day) などの総称。

表 3-1. 発生割合 (リスク) と罹患率の比較

性質	発生割合	罹患率
最小値	0	0
最大値	1	無限大 ( $\infty$ )
単位 (次元)	無し	時間の逆数
解釈	確率	待ち時間の逆数

waiting time

待ち時間。定常的な条件で、時間とともに率が変化しないような状況なら、発生率の逆数は、あるイベントが起こるまでの平均的な時間に等しく、この時間を待ち時間と呼ぶ。

【コラム】鶏と卵

古い謎々が問う。「もし一羽と半分の鶏が一日と半分の間に一個と半分の卵を産んだら、一羽の鶏は、一日に何個の卵を産むか？」この謎は率の問題である。問題内容は、「鶏-日当たりの卵で表現した産卵率は？」を問うものである。答えを得るには、分子を卵数、分母を鶏-日数とした率を計算すればいい。 $1.5 \text{ 卵} / (1.5 \text{ 鶏} \cdot 1.5 \text{ 日}) = 1.5 \text{ 卵} / 2.25 \text{ 鶏日}$ となる。これを計算すると、鶏-日当たり  $2/3$  個の産卵率となるので、謎々の答えは  $2/3$  である。

### 1.3 リスクと罹患率の関係

解釈はリスクの方が直截的なので、罹患率をリスクに変換すると便利なおことがある。最も単純な変換式は、

$$\text{リスク} = \text{罹患率} \times \text{時間} \quad [3-1]$$

である。次元を確認すると両辺とも無次元になっている。次元を確認するときは、値の範囲も確認すると役に立つ。

計算例 2 つ。(1) 1 万人の集団で 1 万人年あたり 8 例の肺がん患者が得られたなら、式 3-1 に当てはめて、この集団の肺がんリスクは  $8/10000$  である (または  $0.0008$  である) といえる。半年で同じ発生率だったなら、リスクはその半分 ( $0.0004$ ) になる。(2) 1000 人の集団で、1000 人年あたり 11 (つまり  $0.011/\text{年}$ ) という死亡率が 20 年間続いたら、式 3-1 から計算されるリスクは  $0.22$  ( $22\%$ ) となるが、この計算は毎年分母が減っていくことを無視しているのだから正しくない。表 3-2

のように1年ずつ計算すると\*1, 20年間で期待死亡は197となり, 20年間のリスクは19.7%となる。

表3-2のような計算を続けると, 図3-3のように, 式3-1から計算されるリスクは, 期待リスクと徐々に乖離していくことがわかる。大きいリスクについては式3-1は使えない。

exponential decay

指数的な壊変。指数的な減衰。死亡を観察するときは死亡が発生するたびに分母も減っていくので, 死亡率が一定なら, 生存しつづけている割合は, 時間とともに指数曲線を描く。このカーブは放射性物質の壊変を記述する曲線と同じである。

発生率は短い時間間隔で調べるほど正確だが, それぞれの間隔について意味のある発生率が得られないと使えない。発生率が時間とともに変化するのは良くあることで, そういう場合は生存時間解析で計算する。

survival analysis

生存時間解析。生存解析。時間とともに変化する発生率についてリスクを計算する方法。死亡だけでなく致死的でないリスクにも使えるが, 元々は死亡に関連したデータから出てきた分析法である。

life-table

生命表。生存時間解析の1つのやり方として, 表3-2のようなもの。関心のある期間を構成している, 連続する時間間隔を通しての生残確率を計算する目的で作る。全期間を通しての生残確率(累積生残確率)は, 各期間の生残確率を累積的に掛け合わせたものになる。表3-3が生命表の簡単な計算例。最終間隔での累積生残確率を1から引くと生涯リスク(lifetime risk)が得られる(誕生から85歳の誕生日までに交通事故で死ぬリスクは $1 - 0.9838 = 1.6\%$ となる。なお, 表3-3の例では全リスクが小さいので分母の逐次減少を考えずに期間ごとの死亡を足しても同じ結果になる)。

(中澤注)なお, 全死因についての生命表は, ゼロ歳平均余命(=平均寿命)を計算するのにも使われる。ここでは詳しく触れられていないが, 特定死因を除去することによる平均寿命の延長分によって, 損失余命という指標も計算できる。

\*1 1年目に11人死んでいない理由は, 1人死んだ時点で999人しか残らないから。1年を $p$ 個に区分して考えると,  $1/p$ 年間で期待生残数は $1000 \times (1 - 0.011/p)$ 人となり, 次の $1/p$ 年間が経った後の期待生残数は $1000 \times (1 - 0.011/p) \times (1 - 0.011/p)$ となり....., という具合に, 最終的に1年経ったときの期待生残数は $1000 \times (1 - 0.011/p)^p$ となる。区分の幅を無限小にする(つまり $p$ を無限大にする)と, この値は0.9890603に近づく(たぶん解析的に解けると思うが, これは $p = 10, 100, 1000, 10000, 100000, 1000000$ として行って収束した数値解)

ハムスターの生命表(架空のもの)

年齢	$m_x$	$q_x$	$l_x$	$d_x$	$L_x$	$T_x$	$e_x$
0	0.5	0.4	100000	40000	80000	213100	2.13
1	0.26	0.23	60000	13805	53097	133100	2.22
2	0.26	0.23	46195	10629	40880	80003	1.73
3	0.5	0.4	35566	14226	28453	39122	1.1
4	1	1	21339	21339	10670	10670	0.5

年齢	$m_x$	$q_x$	$l_x$	$d_x$	$L_x$	$T_x$	$e_x$
0	$m_0$	$q_0 = m_0 / (1 + 2/m_0)$	$l_0 = 100000$	$d_0 = l_0 * q_0$	$L_0 = l_0 - d_0 / 2$	$T_0 = L_0 + T_1$	$e_0 = T_0 / l_0$
1	$m_1$	$q_1 = m_1 / (1 + 2/m_1)$	$l_1 = l_0 - d_0$	$d_1 = l_1 * q_1$	$L_1 = l_1 - d_1 / 2$	$T_1 = L_1 + T_2$	$e_1 = T_1 / l_1$
2	$m_2$	$q_2 = m_2 / (1 + 2/m_2)$	$l_2 = l_1 - d_1$	$d_2 = l_2 * q_2$	$L_2 = l_2 - d_2 / 2$	$T_2 = L_2 + T_3$	$e_2 = T_2 / l_2$
3	$m_3$	$q_3 = m_3 / (1 + 2/m_3)$	$l_3 = l_2 - d_2$	$d_3 = l_3 * q_3$	$L_3 = l_3 - d_3 / 2$	$T_3 = L_3 + T_4$	$e_3 = T_3 / l_3$
4	$m_4 = 1$	$q_4 = 1$	$l_4 = l_3 - d_3$	$d_4 = l_4 * q_4$	$L_4 = l_4 - d_4 / 2$	$T_4 = L_4$	$e_4 = T_4 / l_4$

#### 1.4 点発生源の流行と伝播する流行

epidemic

流行。疾病の異常に高い発生をいう。どの程度だと「異常に高い」のかは状況によって異なる。

outbreak

勃発。突然の大発生。普段はないかほとんどない疾病の発生が突然増えることをいう。

point-source epidemic

点発生源の流行。因果的な病因への単独の曝露から起こった流行。例えば汚染された食事を出されたレストランの顧客の食中毒流行とか。

propagated epidemic

伝播する流行。増殖する流行。インフルエンザの流行のように、因果的な病因それ自体が集団内で伝播されるときに起こるものをいう。

## 1.5 有病割合

prevalence

有病割合。prevalence proportion と書かれることもある<sup>a</sup>。原則として一時点での、総数に対する患者の割合。無次元（人／人なので単位がない）。一時点でのということを明示するには、point prevalence という。急性感染症で prevalence 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。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：高血圧や高コレステロール血症は prevalence が高いので、行政的対策を打つのにいくらかかるかの算定が必要）。その集団における疾病の負荷を示す指標である。

<sup>a</sup> 中澤注：疫学辞典では、prevalence を有病数として分子のみを指すものとし、prevalence “rate” と同じ意味で使うこともある、と記載し、prevalence “rate” の項で、これは率でなく有病割合である、と記載しているが、最近の本や論文では Rothman や Greenland の厳密な考え方がだいぶ広まったので、prevalence rate などと書く人は減ったし、prevalence は一般に prevalence proportion 以外の意味では用いられなくなった。

steady state

定常状態。疾病の発生率と持続期間が時間が経っても安定している状態をいう。

prevalence odds

有病割合オッズ。有病割合のオッズ。

odds

オッズ。一般に何らかの割合をとるときはいつも、それ以外の割合に対する比を計算することができて、それを、その割合のオッズと呼ぶ。競馬で、ある馬のオッズが3対1ということは、その馬が勝つ確率が0.75であることを意味する。勝たない確率0.25に対する比をとると $0.75/0.25=3$ となるので、通常これを3対1（3 to 1）と書く（日本では3倍という）。割合が小さいとき、分母が1に近づくので、オッズはその割合自体と近い値になる。

【コラム】特性の有病割合

有病割合は状態を測るので、集団中の疾病よりも特性あるいは条件の状態を記述するのに用いることが多い。例えば、喫煙習慣をもつ人の割合を喫煙割合（喫煙の有病割合）として記述する。何らかの病因に曝露した割合は、曝露割合と書くことが多い。

## 2 因果的効果の尺度

### 2.1 反事実という理念

counterfactual

反事実。ある病因の疾病への因果的な効果を測定するには、理想的には、その病因への曝露がないことだけが異なっていて、それ以外の条件はすべて同じ人々に対して、その病因に曝露した人々が、その疾病を何倍発生しやすいかを調べるべき。けれども、人は一人一人いろいろな点で異なっているので、ある個人の疾病発生を、ある要因に曝露した場合としなかった場合で比べることは不可能。この理想的な比較条件は現実にはありえないという意味で、これを反事実と呼ぶ。

なお、このことは、個人についての因果推論が不可能であることを示すものでもある。

crossover study

クロスオーバー研究。クロスオーバー法。反事実が理想的ならば、個人個人について、曝露した場合と曝露しなかった場合の両方をペアとして実験的に繰り返してやればいい。個人が一つの試験群から別の群へ経時的に渡り歩くという意味で、こういうデザインをクロスオーバー研究と呼ぶ。

### 2.2 効果の尺度

risk difference

リスク差。曝露群と非曝露群の間の、発生割合（あるいはリスク）の差。

incidence rate difference

発生率差。曝露群と非曝露群の間の、発生率の差。

risk ratio

リスク比。曝露群のリスクの、非曝露群のリスクに対する比。相対的な影響は、リスク比から1を引いたものとなる。

表 3-4. 絶対的な及び相対的な効果の尺度の比較

尺度	数値の範囲	次元
リスク差	$[-1, +1]$	なし
リスク比	$[0, \infty]$	なし
発生率差 (罹患率差)	$[-\infty, \infty]$	時間の逆数
発生率比 (罹患率比)	$[0, \infty]$	なし

【コラム】丸め：何桁表示すべきか？

どこまでの精度が必要かが問題になることは多い。結論としては、一般的な規則は定めがたいといえる。例えば小数点以下一桁までと決めた場合、4.1 ならいいけれど、リスク比のような大抵 1 より小さい値を出したときには 0.7 とか 0.8 とかしか表現できず、丸め誤差が大きくなる。では意味のある桁長 (有効桁長) を表示することにしたらどうか？ 4.1 と 0.83 は同じ正確さをもつので、先のルールよりはずっとましに見えるが、2 桁の精度しかもたせないと、1 付近にかたまりがちな比で表される指標値について、0.98 と 0.99 は区別できるのに 1.00 と 1.01 は区別できなくなるという矛盾が生じる。適切な判断ができればルールは要らない。注意すべきことは、計算途中ではできるだけ丸めない方がいいこと、1.25 のリスク比を 1.2 や 1.3 に丸めるのは効果を 20% も歪めること、丸めが常に四捨五入で行われるのではないこと (例えば最も近い偶数に丸められる場合もある)。

(参考) 戸川隼人「計算機のための誤差解析の基礎」(サイエンス社)

## 2.3 リスク比と率比の関係

instantaneous risk ratio

瞬間のリスク比。発生率比 (incidence rate ratio) と同義。しばしばたんに率比 (rate ratio) とも呼ばれる。

relative risk

相対危険。罹患率比を指すこともあればリスク比を指すこともあるが、一般に相対危険はリスク比と同義と捉えるべき。相対危険よりも、罹患率比とかリスク比という用語の方が望ましい。

#### 【コラム】リスクがリスクを意味しない場合

effect

効果。リスク比は効果の指標の一つである。effect と書くべきところに risk と書いてしまう人もいる。アスベスト曝露からの肺がんの2つのリスク比について、若年成人が5.0、高齢の成人が2.5という研究報告がなされた場合、これらの効果を示す値が、「アスベスト曝露からの肺がんのリスクが高齢者では若年成人ほど大きくない」と記述されることがあるが、正しくない。実際は、曝露群と非曝露群のリスク差は高齢者の方が若年成人より大きいのに、肺がんのリスクが加齢とともに急速に上昇するから、高齢者では分母が大きくなって相対的にリスク比が小さくなる。これはリスク比と書かれるべき、あるいはより一般的な用語として効果と書かれるべきところが、リスクとなっているのが間違っている。リスク比となっていれば正しかった。

## 2.4 寄与分画

attributable fraction

寄与する部分。寄与分画。寄与割合 (attributable proportion) ,病因分画 (etiologic fraction) ,寄与危険 (attributable risk) と同義。一定の集団において、ある因子への曝露の結果、ある疾病が発生したとする。このとき寄与分画とは、その曝露が除去された場合に全集団の罹患率が減少するであろうと思われる割合。