

測定と疫学調査の基礎知識

(Basic knowledge of measurement and epidemiological research)

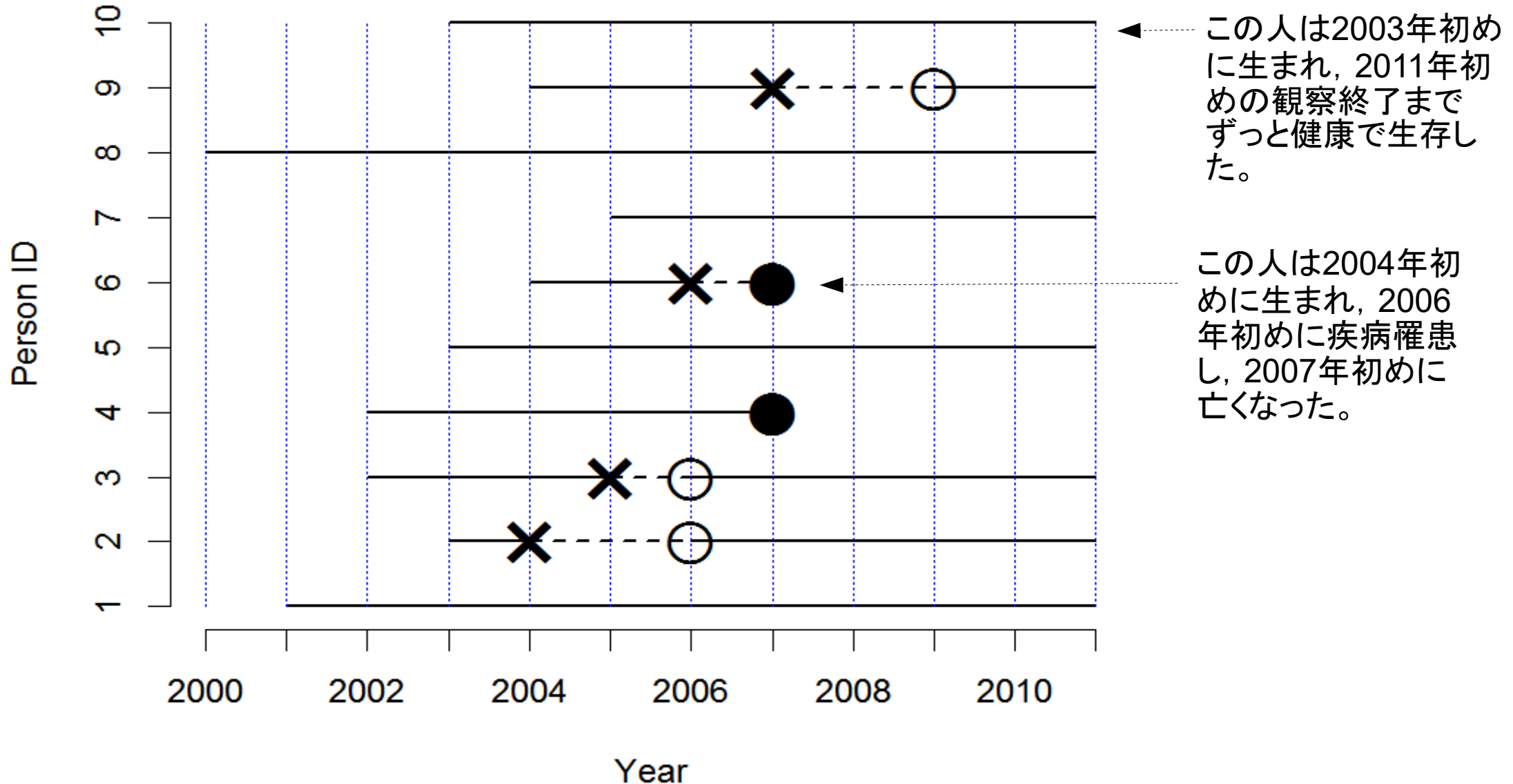
- Validity (妥当性)
 - 測りたいものを正しく測れていること
 - (例) ELISAで抗体の特異性が低いと測定の妥当性が低い
- Accuracy (正しさ, 正確性)
 - バイアス(系統誤差)が小さいこと
 - (例) 測定機器のゼロ点調整に失敗すると正確さが損なわれる
- Precision (精度)
 - 確率的な誤差(ランダムな誤差)が小さいこと
 - 信頼区間が狭い, CV(変動係数)が小さいこととも同義
 - 感度の低い測定は精度が低い

典型的な疫学調査の結果として

- クロス集計表を作る
 - 2つのカテゴリ変数間の関連性の程度を評価
- 疫学調査では、「疾病／非疾病」と「曝露／非曝露」の関連性を調べる
 - 疾病量をどのように把握（評価）するか
 - 断面研究や症例対照研究→prevalenceやodds
 - コホート研究→riskやincidence rate
 - 疾病と曝露の関連性の程度【=effect（効果）】をどのように評価するか
 - difference（差）または ratio（比）
 - p値が有意水準（例えば5%）より小さいことの意味
 - 偶然ではない関連や差がある=統計学的に「有意」（statistically significant）
 - 有意でないとき=何も言えない（差がないとか関連がないとか言えたのではない）
 - 程度の情報を「有意」かどうかという2値情報にしてしまう欠点
 - 信頼区間を提示する方が情報量が多いので最近はそのほうが推奨

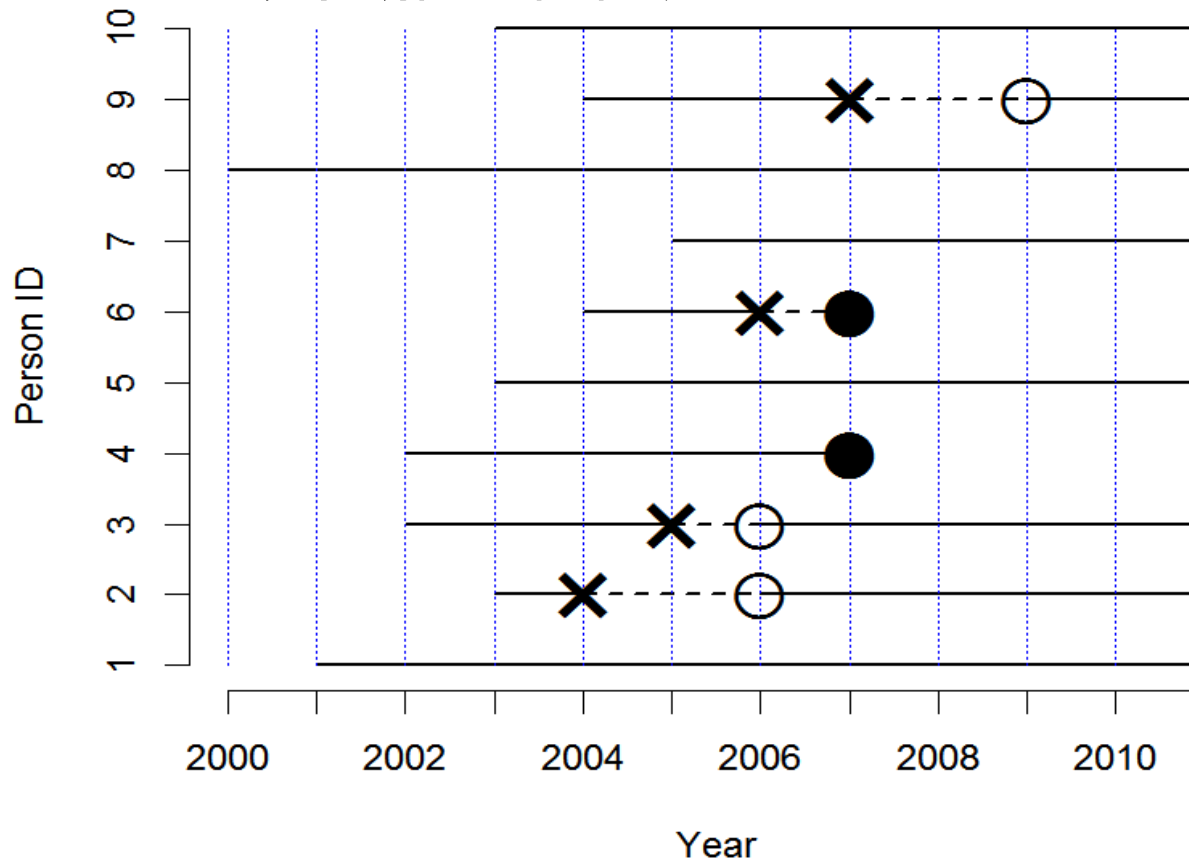
疾病量の把握

- 疾病発生の実際の状態を考えてみよう。下図では、横線のそれぞれが1人の個人を意味し、実線は健康、破線は疾病である状態を示す。×は疾病罹患、○は治癒、●は死亡を意味する。横軸は観察年を意味する。



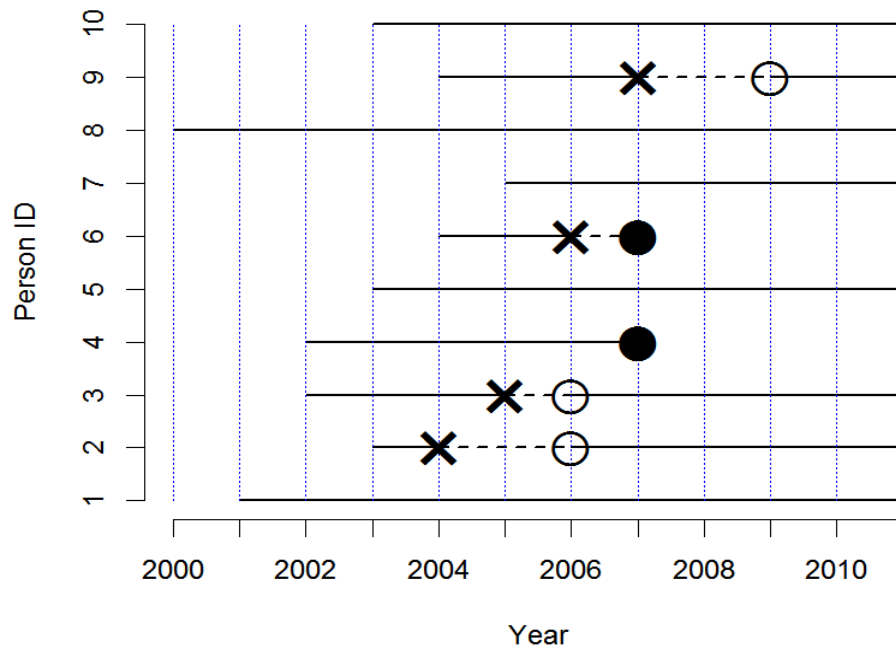
リスク(risk, 累積罹患率ともいう)

- 2011年初めに生き残っている8人の子供について親に過去の罹患について思い出してもらい, 3人について疾病罹患が報告されたなら, この11年間のリスクは $3/8=0.375$ と推定される←研究は簡単で安価にできるが, 既に亡くなった子供を聞き逃してしまう(2000年から毎年あるいはイベントが起こるたびに記録されているデータが存在すれば後ろ向きコホート研究も可能だが稀)
- 2000年から11年間のコホート研究をして, 10人の子供のうち疾病罹患は4人が経験したので, 11年間のリスクは0.4となる。ただし生後1年間に同じ病気に罹るリスクは $1/10=0.1$ ←リスクは観察期間に依存する



罹患率(incidence rate)

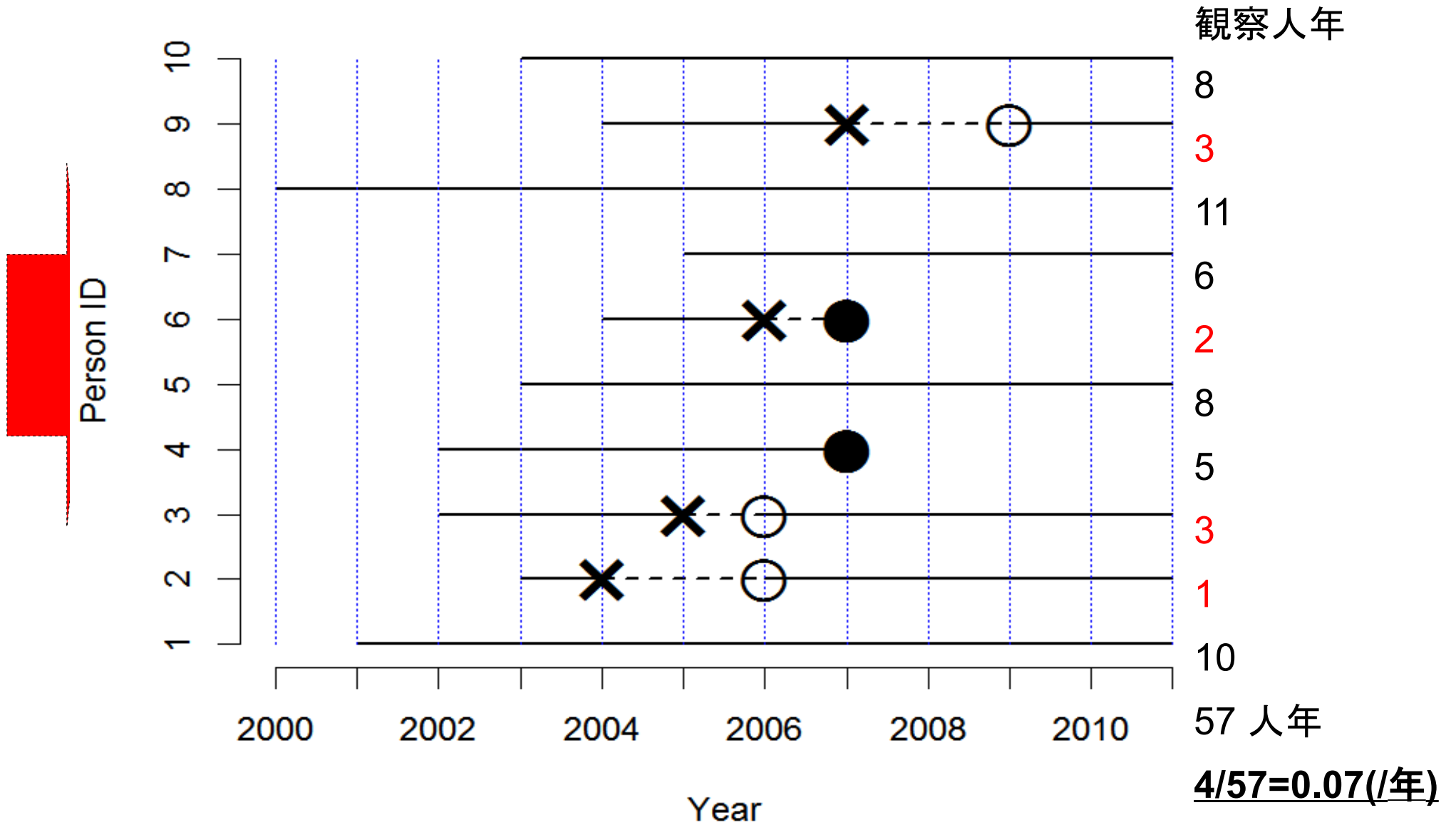
- 2000年からコホート研究すれば全人年データを観察できる
- 注目している疾病が一生に一度しか罹らないものなら、一度罹患した人は感受性を失うので、患者はリスク人口(population at risk)から除去される。感受性のある観察期間の合計人年を分母、疾病発生数を分子にした値は「罹患率」(incidence rate)となる。1/年という次元をもつ。
- 何度も罹患する疾病について、集団の罹患率を求めるには、年央のリスク人口を分母、その年の疾病発生数を分子とすると、罹患率が得られる(通常、10万人年当たりで計算する)。



左の図で罹患率は
どうなる？

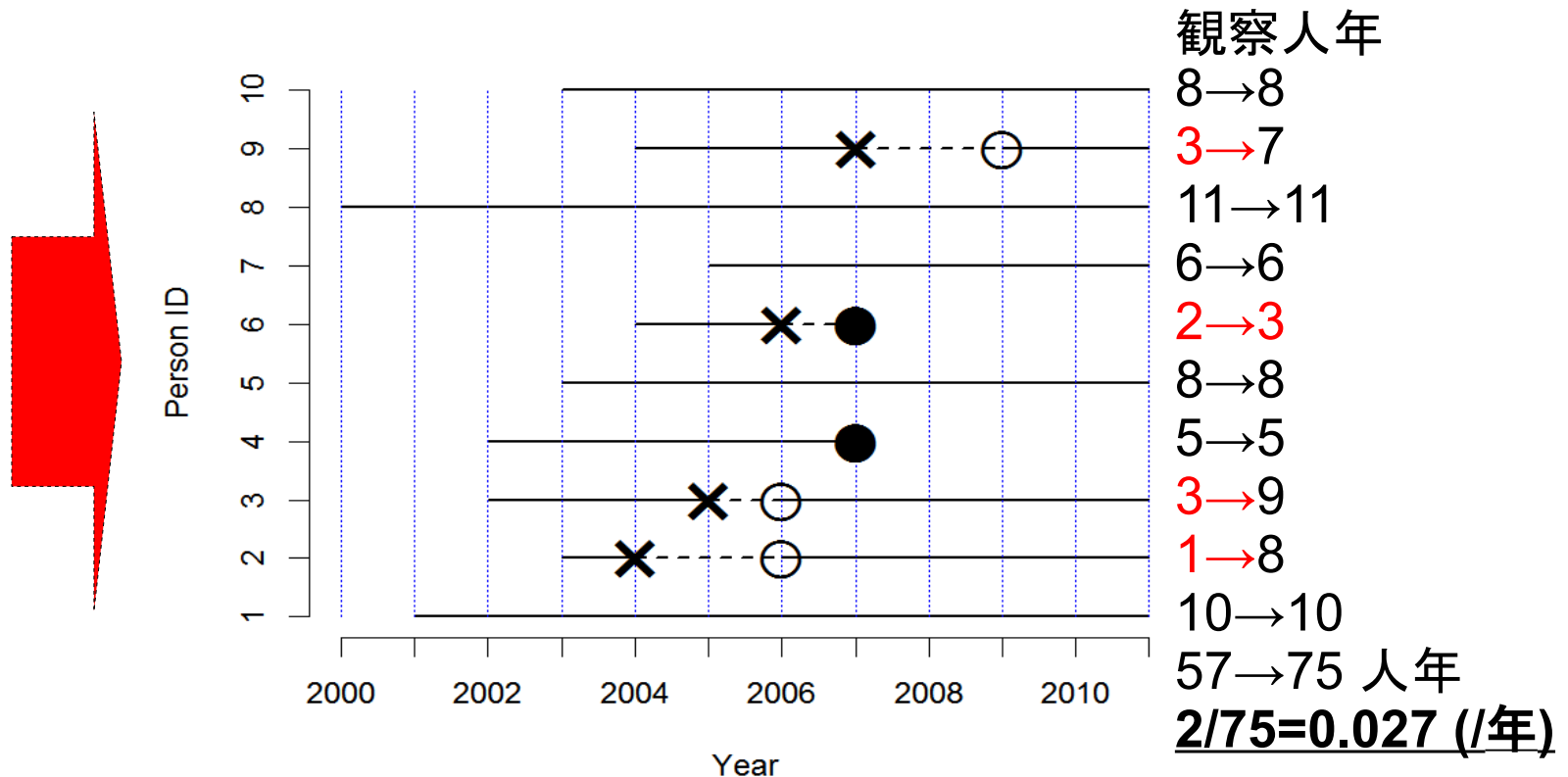
↓
解答は次ページ

コホート研究による罹患率の計算例



死亡率 (mortality rate)

- 観察のエンドポイントを疾病発生から死亡に変えると、罹患率の代わりに死亡率が計算できる→下図のように0.027/年となる
- 死亡は一生に一度しか罹らない疾病と同様に扱う
- 大集団についての指標としては、年間死亡数を年央人口で割るとその年の死亡率が得られる(通常、1000人年または10万人年当たりで表す)
→この例で計算すると、2007年に0.2/年、他の年は0となる



曝露と疾病の関係

= 曝露と非曝露の間で疾病量を比較する

- 典型的な比較方法
 - 差(絶対比較)
 - 比(相対比較)
 - 両方ともそれぞれ意味がある
- 比較する疾病量による違い
 - リスクの場合→リスク差またはリスク比
 - 罹患率の場合→罹患率差または罹患率比
 - 死亡率の場合→死亡率差または死亡率比
 - 有病割合の場合→オッズ比(有病割合の差を比べることはあまりない)

絶対比較: リスク差または率差 = 寄与危険 (超過危険ともいう)

- (架空の例) 高圧線の近くの住民10万人(曝露群)を5年間追跡調査して, 毎年2人ずつ白血病患者が見つかったとする。高圧線から離れたところの住民10万人(非曝露群)を同じ期間追跡調査して, 毎年1人ずつ白血病患者が見つかったとする。この人たちに居住地以外の(教育歴とか年収とか人種とか気候とか)違いはないものとする。
- リスク差 = $10/100000 - 5/100000 = 5/100000 (=5e-5)$
- 罹患率差 =
 $10/(100000+99998+99996+99994+99992)$
 $- 5/(100000+99999+99998+99997+99996)$
 $\approx 0.0000100006 (/year)$
- 罹患率自体が非常に小さいので, 差は小さく見える
- Rでは, library(fmsb)を使って信頼区間やp値も出せる
`riskdifference(10, 5, 100000, 100000)`
`ratedifference(10, 5, 499980, 499990)`

相対比較(1): リスク比と罹患率比

- 同じ例
- リスク比 = $(10/100000)/(5/100000)=2$
- 罹患率比 =
$$\frac{10/(100000+99998+99996+99994+99992)}{5/(100000+99999+99998+99997+99996)}$$
$$\approx 2$$
- どちらの値も, 「高圧線の近くに住むと白血病のリスクが約2倍になる」ことを意味する
- 統計学的に有意?
 - 「比が1」という帰無仮説を検定する
 - R (EZR), SAS, JMPのようなソフトを使えば簡単→p値は0.2より少し小さい程度なので有意水準5%では有意でない
 - `library(fmsb); riskratio(10, 5, 100000, 100000);
rateratio(10, 5, 499980, 499990)`

相対比較(2): オッズ比

- 断面研究では有病割合が疾病量の指標になる。しかし有病割合そのものの差や比は無意味なので評価しない。代わりに属性あり群の疾病オッズを属性無し群の疾病オッズで割ったオッズ比を計算する
- 症例対照研究では、オッズ比は症例の曝露オッズ(曝露有りの人数の曝露無しの人
数に対する比)を対照の曝露オッズで割った値になる
 - 曝露群と非曝露群を合わせて200000人、5年間追跡しても、全部で15人しか発症しないような稀な疾患(リスクが低すぎる疾患)は、コホート研究で研究するのは効率が悪い(要因曝露の影響を直接示せる利点があるので、必要ならやらざるを得ないが)(前掲の例でも5%有意にならなかった)
 - 症例対照研究では、例えば、ある病院に来院した白血病患者100人を症例群、同じ病院に来院した外傷患者200人を対照群として、それぞれに居住地の近くに高圧線があるか無いかを尋ねる。仮に症例群でも対照群でも20人ずつが高圧線の近くに住んでいた場合、

	白血病	外傷
高圧線近い	20	20
高圧線遠い	80	180

というクロス集計表ができて、オッズ比は $(20/80)/(20/180)=2.25$
Rでfisher.test(matrix(c(20, 80, 20, 180),2))とすると5%有意
library(fmsb); oddsratio(20, 80, 20, 180)でも良い