

罹患率・死亡率の計算、年齢調整の方法

地域がん登録によって得られた罹患データは、罹患数・罹患率として計算される。また、計算した罹患率を、他のがん登録実施地域や登録地域内(市町村など)、過去の値と比較することによって、その地域のがん罹患状況の特徴が明らかになる。一方、がん死亡は、がん罹患とそれに続く治療がもたらす最終的な結果あり、罹患と同時に死亡についても比較観察することが重要である。

1. 罹患率・年齢調整罹患率

(1) 罹患率 (incidence rate)

がん罹患数とは、ある集団で新たにがんと診断された(再発を含まない)人数のことである。がん罹患率とは、ある期間の罹患数のことである。通常は1年間の10万人あたりの罹患数で表現される。つまり、x年のある地域の10万人あたりのがん罹患率は、

$$\frac{x \text{ 年に新たにがんを診断された人数}}{x \text{ 年の観察人数(人口)} \times 100000}$$

である。罹患率は、個人のある一時点におけるがん罹患のリスクを表す。

(2) 観察人数 (population at risk)

地域がん登録で罹患率を計算する際の分母となる観察人数とは、罹患数を実測した登録対象地域の人口であり、その地域の年央人口を1年間観察したとして分母とする。登録対象に外国人を含まない場合は、日本人人口を用い、含む場合は総人口とする。通常は分子となる罹患数に在日外国人を含むので、総人口を用いる。

一般的には、ある病気の罹患率を計算する場合には、新規罹患症例が発生した時点で、分母からその症例を除く。しかし、登録地域における新規がん罹患数は、人口に比べて小さく、がん罹患患者を分母から除外しない影響は無視できるほど小さいため、年央人口を1年間観察したとみなしてこれを分母とする。

(3) 粗罹患率 (crude rate) と年齢階級別罹患率 (age-specific rates)

全罹患数とその年の人口で除したものを、粗罹患率という。

年齢階級別の罹患数を対応する年齢階級の人口で除すと、年齢階級別罹患率となる。年齢の区分は、0～4、5～9、10～14、…、80～84、85+歳の5歳区分18階級とすることが多い

が、集計の目的に応じて区分を変えることもある(0歳と1-4歳を別々に計算する場合や、15歳区分とする場合など)。

(4) 年齢調整罹患率 (age-standardized rates)

地域がん登録で罹患率を計算する目的のひとつは、得られた罹患率を他地域や国全体、あるいは、他国の罹患率と比較すること、年次推移の観察を行うことである。比較対象間の人口年齢構成が異なっている場合、年齢の違いを考慮しない粗罹患率による比較では解釈が困難である。例えば、異なる二つの地域の年齢階級別罹患率が全く同じ場合でも、がん罹患率が高い高齢層に人口年齢構成が偏っているほど、粗罹患率は大きくなる。そこで、他の地域のがん罹患率と比較する時や、同じ地域でがん罹患率の動向を観察する時には、人口構成を調整した(人口年齢構成の違いを取り除いた)罹患率、つまり年齢調整罹患率を用いて比較を行う。

ただし、年齢調整罹患率は、比較対象地域が多い場合には簡便で解釈しやすいが、あくまでも要約値であるので、詳細な比較を行う場合には、年齢階級別罹患率を観察すべきである。

年齢調整罹患率には、計算したい地域の人口の構成が基準(標準)人口(standard population)と同じであると仮定して算出する直接法(direct method)と、基準(標準)人口集団での年齢階級別罹患率を用いて計算する間接法(indirect method)がある。

1) 直接法

比較する対象間で年齢構成に偏りがある場合、標準とする集団の人口構成と同一であると仮定した場合の仮の率を計算して比較する。標準とする集団を基準(標準)人口という。

直接法で年齢調整罹患率を計算する際の基準(標準)人口は、比較する目的によって選ぶ。地域がん登録では、国内の他地域との比較や年次推移の観察には「昭和60年モデル人口」を、世界各国との比較には「世界人口」を用いて年齢調整罹患率を比較する。年齢調整罹患率は人口10万対で表される。

年齢調整罹患率は、標準人口の年齢分布で重み付けした仮の値であるため、単独値で大小の議論を行うことには意味がない。

(基準(標準)人口)

世界人口

国際比較を行うときに用いる基準(標準)人口。IARC 刊行の 5 大陸のがん罹患(Cancer Incidence in Five Continents)の年齢調整罹患率の算出に用いられている。

昭和 60 年モデル人口

国内比較を行うときに用いる基準(標準)人口。世界人口よりも高齢者に重みがかかっており、日本の現状をより反映した年齢調整罹患率が得られる。

| 年齢 | 昭和60年 モデル人口 | 世界人口 |
|-------|----------------|--------|
| 0-4 | 8180000 | 12000 |
| 5-9 | 8338000 | 10000 |
| 10-14 | 8497000 | 9000 |
| 15-19 | 8655000 | 9000 |
| 20-24 | 8814000 | 8000 |
| 25-29 | 8972000 | 8000 |
| 30-34 | 9130000 | 6000 |
| 35-59 | 9289000 | 6000 |
| 40-44 | 9400000 | 6000 |
| 45-49 | 8651000 | 6000 |
| 50-54 | 7616000 | 5000 |
| 55-59 | 6581000 | 4000 |
| 60-64 | 5546000 | 4000 |
| 65-69 | 4511000 | 3000 |
| 70-74 | 3476000 | 2000 |
| 75-79 | 2441000 | 1000 |
| 80-84 | 1406000 | 500 |
| 85歳以上 | 784000 | 500 |
| 総数 | 120287000 | 100000 |

(図 1) 基準(標準)人口: 昭和 60 年モデル人口と世界人口

2) 間接法

間接法により得られる値は、正確には年齢調整罹患率ではなく、期待値と観測値の比である。

対象とする地域(例えば市町村)の年齢階級別罹患率が、比較しようとする集団(例えば県全体)の年齢階級別罹患率と同じと仮定した場合の罹患数(期待罹患数)を計算し、実際に

観察された罹患数(観察罹患数)との比[標準化罹患比(SIR): standardized incidence rate]を求める。対象とする地域の年齢階級別罹患率がわからなくとも、年齢階級別人口と全年齢階級での観察罹患数が得られている場合には、標準化罹患比を計算することができる。

この方法は、人口規模の小さい集団(市町村や医療圏など)の罹患を、全県など基準とする集団と比較したい場合に用いることが多い。人口規模の小さい集団で年齢階級別罹患率を求めると偶然変動により値が安定せず、偏った値になる可能性が高いからである。

標準化罹患比が 1 の場合は、期待罹患数と同じ、つまり比較集団と同じ、1 より大きい場合は比較集団よりもがん罹患が多く、1 より小さい場合は、がん罹患が少ないことを表す。

間接法による標準化のための期待値の計算は、対象集団の人口構成に依存しており、重み付けが対象集団間で異なる。従って、対象集団の標準化罹患比は、基準とする集団と比較はできるが、対象集団同士の比較は厳密には出来ない。対象集団間での比較は、対象集団と比較集団の死亡率の比が年齢に依存しないと仮定のもとで可能である。

2. 罹患率・年齢調整罹患率の計算

(1) 年齢階級別罹患率と粗罹患率

基準(標準)集団および比較対象集団のある期間における年齢階級別人口、がん罹患数、がん罹患率(罹患数/年齢階級別人口)を表 1 の様に表す。

このとき、

$$\text{粗罹患率} = \frac{r}{n}$$

i 番目の年齢階級の罹患率(p_i)は、

$$p_i = \frac{r_i}{n_i}$$

| 年齢階級 | 基準(標準)集団 | | | 比較対象集団 | | |
|------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | ① 人口 | ② 罹患数 | ③ 罹患率 ②/① | ④ 人口 (観察人数) | ⑤ 罹患数 | ⑥ 罹患率 ⑤/④ |
| 1 | N_1 | R_1 | P_1 | n_1 | r_1 | p_1 |
| 2 | N_2 | R_2 | P_2 | n_2 | r_2 | p_2 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| i | N_i | R_i | P_i | n_i | r_i | p_i |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| k | N_k | R_k | P_k | n_k | r_k | p_k |
| 計 | $N = \sum_{i=1}^k N_i$ | $R = \sum_{i=1}^k R_i$ | $P = \sum_{i=1}^k P_i$ | $n = \sum_{i=1}^k n_i$ | $r = \sum_{i=1}^k r_i$ | $p = \sum_{i=1}^k p_i$ |

(表 1) 年齢階級、基準(標準)集団と比較対象集団の人口、罹患数、罹患率

(2) 年齢調整罹患率の計算と信頼区間

年齢調整罹患率の信頼区間[confidence interval (CI)]は、次式で表される。

$$ASR \pm Z_{\alpha/2} \times [SE(ASR)]$$

$SE(ASR)$: ASR の標準誤差 (Standard error)

Z : Z 値 (Standard normal deviate)

95%信頼区間を求める場合、

$$Z_{\alpha/2} = 1.96、$$

99%信頼区間を求める場合、

$$Z_{\alpha/2} = 2.58 \text{ となる。}$$

以下、直接年齢調整罹患率、および、標準誤差と信頼区間の計算を示す。

1) 直接法

年齢調整罹患率は、比較対象集団の人口構成を基準(標準)人口にあてはめたものである。また、 N_i で重み付けした p_i の平均である、とも解釈できる。

年齢調整罹患率(ASR)

$$= \frac{\sum_{i=1}^k N_i p_i}{\sum_{i=1}^k N_i}$$

ASR が 2 項分布に従うとすると ASR の分散 [$Var(ASR)$] は、

$$Var(ASR) = \frac{\sum_{i=1}^k \left(\frac{N_i^2 p_i q_i}{n_i} \right)}{\left(\sum_{i=1}^k N_i \right)^2}$$

($q_i = 1 - p_i$)

p_i が十分に小のとき(がん罹患やがん死亡の場合)、次式に置き換えが可能である。

$$Var(ASR) = \frac{\sum_{i=1}^k \left(\frac{N_i^2 p_i q_i}{n_i} \right)}{\left(\sum_{i=1}^k N_i \right)^2}$$

つまり、

$$SE(ASR) = \sqrt{Var(ASR)} = \frac{\sqrt{\sum_{i=1}^k \left(\frac{N_i^2 p_i q_i}{n_i} \right)}}{\sum_{i=1}^k N_i}$$

2) 間接法

比較対象集団の全罹患数は、

$$r = \sum_{i=1}^k r_i$$

比較対象集団の各年齢階級別罹患率が基準(標準)集団と同じと過程した場合の期待罹患数は、

$$\sum_{i=1}^k n_i P_i$$

標準化比 (SIR: Standard incidence ratio) は

$$M = \frac{\sum_{i=1}^k r_i}{\sum_{i=1}^k n_i P_i}$$

死亡データの場合、標準化死亡比 (SMR: Standard mortality ratio) という。一般的に、SIR および SMR は、パーセントで表す

$$SIR / SMR = M \times 100 \quad (\%)$$

標準化比 M の分散は、

$$\text{Var}(M) = \text{Var} \left(\frac{\sum_{i=1}^k r_i}{\sum_{i=1}^k n_i P_i} \right)$$

ここで、2 項目分布を仮定すると、年齢別罹患数の分散は、

$$\text{Var}(r_i) = n_i p_i q_i \\ (q_i = 1 - p_i)$$

したがって、

$$\text{Var}(M) = \frac{\sum_{i=1}^k n_i p_i q_i}{\left(\sum_{i=1}^k n_i P_i \right)^2}$$

p_i が十分に小さいとき

$$\begin{aligned} & \approx \frac{\sum_{i=1}^k n_i p_i}{\left(\sum_{i=1}^k n_i P_i \right)^2} \\ & \approx \frac{\sum_{i=1}^k r_i}{\left(\sum_{i=1}^k n_i P_i \right)^2} \end{aligned}$$

つまり、

$$SE(M) = \frac{\sqrt{\sum_{i=1}^k r_i}}{\sum_{i=1}^k n_i P_i}$$

3. 累積罹患率 (cumulative incidence rates) と累積罹患リスク (cumulative incidence risk)

累積リスクとは、他の疾患で死亡しないと仮定した場合の、ある年齢区間 (通常 0~74 歳)

において個人ががん罹患する確率である。

累積罹患率は、1 歳年齢階級別罹患率の合計値であり、年齢階級別人口が同じ場合の直接的な年齢調整罹患率であると解釈できる。また、累積罹患率はその値が十分小さいとき (例えばがんの罹患率) は、累積罹患リスクとほぼ同様の値となる。累積罹患率は、直接法と間接法以外の年齢調整の方法として Day (1987) によって提唱され、5 大陸のがん罹患でも第 4 巻から値が掲載されている。

累積罹患率 (Cum rate) は、次式より計算する。

$t_i = i$ 番目の年齢階級の幅 (t 年) のとき、

$$\text{Cum.risk} = \sum_{i=1}^k p_i t_i$$

(例) 0 ~ 74 歳の累積罹患数 :

$$\text{Cum.rate} = \sum_{i=1}^{15} 5r_i$$

このとき、累積罹患リスク (Cum.risk) は、

$$\text{Cum.risk} = 1 - \exp(-\text{cum.rate})$$

Cum.rate が十分小さいとき、

$$\text{Cum.rate} \approx \text{Cum.risk} \quad \text{となる}$$

罹患率は、個人のある一時点におけるがん罹患へのリスクを表す「率」であるのに対し、累積罹患率は、個人が一定の年齢内にごんを患う危険度を表す「割合」であり罹患する確率である。通常パーセンテージで表す。

累積罹患率は、(1) 計算に基準 (標準) 人口を選択する必要がない、つまり基準 (標準) 人口による重み付けの影響を受けない、(2) 異なる年齢階級の累積罹患率を求める場合は率同士を足すことができる (0~74 歳の累積罹患率 = 0~39 歳の累積罹患率 + 40~74 歳の累積罹患率)、(3) $1 - \exp(-\text{累積罹患率})$ の式により、簡単に累積罹患リスクが求められる、という利点がある。

4. 死亡率・年齢調整死亡率

がん罹患は、がんという事象の発生率である。死亡も同様だがんによる死亡という事象の発生率である。したがって、がん死亡率 (mortality rates) ・年齢調整死亡率 (age-standardized mortality rates) ・累積死亡率 (cumulative mortality rates) ・累積死亡リスク (cumulative mortality risk) の計算の方法はがん罹患率・年齢調整罹患率と同様である。