

# 研究員 の眼

## バークソンのバイアス 合併症の起こりやすさは本当に異なるのか？

保険研究部 主席研究員 篠原 拓也  
(03)3512-1823 tshino@nli-research.co.jp

病気の発生動向を研究する学問分野として、疫学がある。ある集団をもとに、病気の発生動向を観測する。そして、得られたデータを統計処理する。その結果、「〇〇の健康状態の人は、◇◇の病気にかかりやすい」といったことを明らかにする。

昨今のデータサイエンスの隆盛を背景に、疫学の分野では、さまざまな統計解析が進められている。そして、そこで得られた結果をもとに、有効な健康増進策を推進しようという動きも出てきている。

ただし、疫学の統計解析には、一筋縄ではいかない点があるので注意が必要だ。たとえば、病気の合併症について、つぎのような分析をしたとする。

### 【病気の合併症に関する分析】

- (1) 病気Aと病気Cの合併症と、病気Bと病気Cの合併症では、一体どちらが起こりやすいのか？ある病院の診療記録をもとに、割り出すことにしたい。
- (2) 病気Aにかかっていた人は620人。そのうち80人は病気Cにもかかっていた。したがって、AとCの合併率は、13% (=80人÷620人)。  
  
一方、病気Bにかかっていた人は240人。そのうち60人は病気Cにもかかっていた。したがって、BとCの合併率は、25% (=60人÷240人)。
- (3) 2つの合併率を比べると、病気Bと病気Cのほうが、病気Aと病気Cよりも高い。そこで、「病気Bと病気Cの合併症のほうが起こりやすい」と結論づけた。

一見すると、この分析には、特段、問題がないようにみえるかもしれない。しかし、実は、大きな問題が潜んでいる。読者のみなさんは、お気づきだろうか？

「ある病院の診療記録」をもとに、合併率を計算していることが問題なのだ。

どんな病気でも、患者がすべて病院に来て診療を受けるわけではない。身体の調子が多少悪くても、しばらく様子を見る人や、市販薬を使って自分で治そう（いわゆるセルフメディケーション）という人がある。特に、急激な痛みなどを伴わない慢性の病気で、病状が徐々に進行していくような場合には、そうした傾向が強くなるだろう。

つまり、病気にかかっているが、病院で受療していない「潜在患者」について、考えてみる必要がある。

先ほどの例で、「病気Aにかかった患者の6割、病気Bにかかった患者の2割、病気Cにかかった患者の5割が、かかってすぐに病院で受療した」としてみよう。なお、最初から合併症として受療するケースは考慮しないことにする。

このような場合、潜在患者も含めて患者全体でみると、合併率はどうなるだろうか。

まず、病気Aの患者からみてみよう。Aにはかかっているが、Cとの合併症ではない患者は、540人（=620人−80人）受療していた。これは、患者全体の6割で、背後には、まだ4割の潜在患者がいる。つまりAにはかかっているが、Cとの合併症ではない患者は、潜在患者を含めると、900人（=540人÷0.6）いることになる。

AとCの合併症の患者はどうか。潜在患者を含めると、全部でX人いるとしよう。そのうちの6割が病気Aで受療し、残りの5割が病気Cで受療する。これを方程式で表すと、

$$X \times 0.6 + (X - X \times 0.6) \times 0.5 = 80 \text{ 人}$$

となる。この方程式を解くと、 $X=100$ 人となる。

つぎに、病気Bの患者をみてみよう。Bにはかかっているが、Cとの合併症ではない患者は、180人（=240人−60人）受療していた。これは、患者全体の2割に過ぎず、背後には、さらに8割の潜在患者がいる。つまりBにはかかっているが、Cとの合併症ではない患者は、潜在患者を含めると、900人（=180人÷0.2）いることになる。

BとCの合併症の患者はどうか。潜在患者を含めると、全部でY人いるとしよう。そのうちの2割が病気Bで受療し、残りの5割が病気Cで受療する。これを方程式で表すと、

$$Y \times 0.2 + (Y - Y \times 0.2) \times 0.5 = 60 \text{ 人}$$

となる。この方程式を解くと、 $Y=100$ 人となる。

まとめると、病気AもBも、それぞれ1000人の患者がいて、病気Cとの合併症の患者が100人ずつ

いる、ということになる。合併率はどちらも 10%で、合併症のかかりやすさは同じという結果だ。

(まとめ)

病気 A の患者の 6 割、病気 B の患者の 2 割、病気 C の患者の 5 割が病院で受療すると仮定した場合

	受療患者+潜在患者	受療患者のみ
病気 A のみに罹患	900 人	540 人
病気 A と C に罹患(合併症)	100 人	80 人
病気 A の患者	1000 人	620 人
<b>病気 A と病気 C の合併率</b>	<b>10%</b>	<b>13%</b>
病気 B のみに罹患	900 人	180 人
病気 B と C に罹患(合併症)	100 人	60 人
病気 B の患者	1000 人	240 人
<b>病気 B と病気 C の合併率</b>	<b>10%</b>	<b>25%</b>
合併率の比較	潜在患者を含めると、合併率は同じ <b>[真相]</b>	受療患者のみでは、病気 B と病気 C の合併率のほうが高い <b>[表面的な見え方]</b>

潜在患者を含めた[真相]は、病気にかかった患者の何割が病院で受療するか、に依存している。この表では、病気 A は 6 割、病気 B は 2 割、病気 C は 5 割としたが、これと別の割合を仮定すれば、計算結果は、また違ったものとなる。

この現象は、アメリカの医師であり統計学者でもあった、ジョセフ・バークソンが 1946 年に公表したことから、「バークソンのバイアス」と呼ばれている。彼は、糖尿病と胆のう炎の併発の可能性について、病院の記録をもとに行う研究を例に、問題点を指摘した。

現在、ICT(情報通信技術)の高まりを背景に、データ管理やさまざまな統計解析のツールが利用可能となっている。機械的にデータを統計処理して、そこから結論を導くこともできる。しかし、いくら正しい統計処理を行っても、もともとのデータ自体がバイアスに汚染されている可能性がある。

疫学にかかわらず、データサイエンスにもとづく、統計の結果を目にするときには、データの取得にバイアスが絡むなどの問題はないのか、という視点を持つことが必要と思われるが、いかがだろうか。

(参考文献)

「医学統計学シリーズ 1 統計学のセンス デザインする視点・データを見る目」丹後俊郎著(朝倉書店, 1998 年)