

放射線生態リスクを評価するための統計学

佐藤 泰憲

Sato Yasunori

(千葉大学大学院医学研究院 グローバル臨床試験学)

1 はじめに

日常生活の中には、災害、事故、病気等の様々なリスクがあり、生命・健康を脅かすものは数多くある。そのため、リスクというと危険というイメージが強いのではないだろうか。我々は、病気になると薬を服薬する人が多いと思われる。「くすり」を逆から読むと「リスク」である。昔から、「毒と薬は紙一重」と言われ、薬はもともと毒であるが、これを注意して上手に利用すれば、病気を治すことができる¹⁾。まさに、「くすり」と「リスク」は表裏一体の関係である。そのため、リスクは危険のみを考えるのではなく、「ベネフィットを得るためにリスクを負う」という考えも必要であろう。

リスクを科学的に評価するためには、何らかの合理的な定義が必要であるが、様々な分野においてそれぞれ関心のある事象や指標が異なっており、統一的な定義をすることは困難である。放射線の生態リスク評価では、「リスク」は放射線被ばくによって生ずる有害な健康影響の大きさを表現するための定量的概念として用いられており、様々な指標が利用されている。これらのリスクはデータに基づいて評価するため、データの取得・提示・活用の各側面で、データの取り扱いが重要である。本稿では、データの提供と利用の注意点・留意点について、統計学的視点から概説する。なお本稿は、科学研究費補助金基盤研究(B)「レギュラトリーサイエンスを導入した放射線教育プログラム開発ならびに教員養成」(研究代表者：杉田克生)の研究成果である「放射線リスク教育ガイドブック」²⁾を加筆したものである。

2 データの集め方

もともと統計というものは、本当に存在している人の集団について、人数や所有物の量等を調べ、数量的にその集団を把握するものであった。集団全体を調べなくても、集団の一部を調べれば全体がかなり把握できることが分かり、集団の一部のみを調べることが行われるようになった(図1)。このときの集団の中で調べられる一部分を集団からの「標本」と呼び、元の集団を「母集団」という³⁾。

筆者らが扱うデータの多くは母集団から抽出された標本である。データを評価する際に、①母集団を規定、②標本の抽出方法が重要なポイントである。母集団の規定においては、標本が母集団の主な特徴を反映するように取得されていなければならない。例えば、定点観測地点の放射線量測定データは、測定機器設置場所周辺の測定時点での大気を母集団とすることが正しいが、測定地点から離れた住民の生活圏の大気を母集団とすることは正しくない。次に、標本の抽出方法は、理想的には、母集団からランダムに選ばれることが望ましい。ランダムとは、母集団のどの部分も標本に選ばれる可能性が同じであることを意味する。

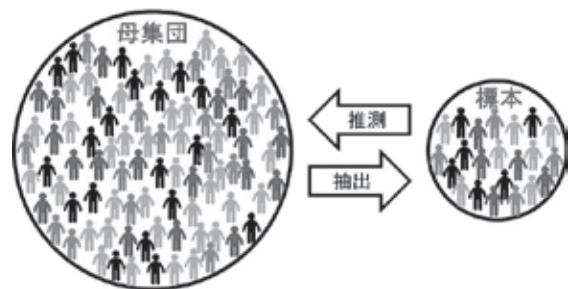


図1 母集団と標本

放射線生態リスクを評価するために、収集されたデータの母集団がどのような集団であるか定義を明確にし、どのようにデータが収集されたかを確認することが重要であろう。

3 交絡

調査研究では、データの取り方・まとめ方によって、全く異なった結論が導かれてしまうことがある。仮想的な事例として、低自然放射線地域（地域①）の住民と高自然放射線地域（地域②）の住民を20年間調査し、死亡割合を比較した。その結果、表1のような結果が得られた。

65歳以上のグループでは、どちらの群も死亡割合は80%であり高い。一方、65歳未満のグループでは、どちらの群も死亡割合は20%であり低い。また、この結果から地域①と地域②との死亡割合は同じである（表2）。

このように結果のまとめ方によって、なぜ結論が変わってしまうのだろうか。図2の結果をよくみると、地域①では、全体の被験者数に対して65歳以上の割合が、地域②よりも高く、地域②では地域①よりも65歳未満の占める割合が高いことが分かる。つまり、65歳以上の高齢者は、20年も追跡調査をすれば寿命等で亡くなるリスクが高いことから、高

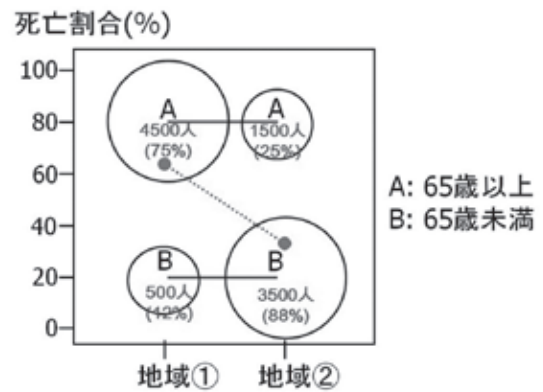


図2 交絡の仕組み

齢者が多い地域ほど死亡割合が高くなる。

このように、「興味ある要因」が「結果」に与える効果を正しく評価できない現象を交絡といい、そのような影響する要因を交絡因子という⁴⁾。この事例では、年齢が交絡因子である。

研究から得られた結果を評価する際に、交絡の存在を見落としてしまうと、誤った結論が導きだされる可能性がある。そのため、交絡因子の存在と交絡の可能性を常に意識して統計解析することが重要である。また、研究開始前に交絡因子の存在が分かっている場合は、交絡を除去するような研究計画を立てることが重要である。

4 統計モデルを利用したリスク評価

我々の生活において、「モデル」という言葉をよく見聞きする。モデルという術語は、分野によって異なる意味を持っているが、その異なる中にもある種の共通性がある。それはある形質に注目しているときには代用品として使えるものという性質である³⁾。統計学で用いるモデルは、確率変数という概念を導入して、測定値の構造を数式で表現したものであり、数学モデル、確率モデルと同様である⁵⁾。

放射線リスク評価の重要な課題の1つは、がんの発症に関して、放射線量や性別、年齢、喫煙、家族歴、被爆歴等の要因を考慮してリスクを数量的に推定することである。そのための方法の1つが、統計モデルを利用したものである。例えば、我が国の原爆被爆者の疫学調査データに基づき、がんによる死亡年齢、被爆時年齢、放射線被ばく線量、性別をパラメータとして、がんの発症リスクを推定するモデ

表1 2つの地域の死亡割合

地域	死亡	生存	合計
地域①	3,250 (65%)	1,750 (35%)	5,000
地域②	1,900 (38%)	3,100 (72%)	5,000
合計	5,150 (52%)	4,850 (48%)	10,000

表2 年齢で層別した結果

地域	年齢	死亡	生存	合計
地域①	65歳以上	3,150 (80%)	1,350 (20%)	4,500
	65歳未満	100 (20%)	400 (80%)	500
地域②	65歳以上	1,200 (80%)	300 (20%)	1,500
	65歳未満	700 (20%)	2,800 (80%)	3,500

ルが提案されている⁶⁾。そのモデルから、被ばくによって増加する年齢別のがん死亡確率や1 Sv 当たりのがん誘発の確率を臓器ごとに算出することができる^{7,8)}。

このように統計モデルを利用すると、要因ごとに定量的にリスクを推定することができるため、非常に有益である。しかし、統計モデルを利用する際に注意すべきことは、モデルは現実と完全に同じではないことである。例えば、上記のがんの発症リスクモデルでは、過剰相対リスク（「被ばくで増加したがん死亡率」の「自然がん死亡率」に対する比）が、時間によらず一定であると仮定している。現実的には、リスクの時間的な変化が一定とは限らないため、このような仮定が妥当なのか吟味した上で統計モデルを利用する必要がある。

医学・疫学研究から得られたデータは、メカニズムや因果関係が不明であることが多い。このような場合に、データの特徴をモデル化して有用な情報を集約することが重要である。また、データの図示を通じて、変数間の関連を見つけ出し、因果関係の有無やデータが特殊な構造になっていないかを確認し、データの発生過程を考慮してモデル化することも重要であろう。

著名な統計学者 George Box は「すべての統計モデルには間違いがあるが、役に立つモデルも存在する」⁹⁾と述べているように、統計モデルが現実に完璧に当てはまることはなく、誤った結論を導かない

ように検証を行い、十分に吟味し統計モデルに基づいた解析結果を受け入れるというのが現実的態度であろう。そのため、統計家・疫学者のみで統計モデル作成やその結果の解釈を行うのではなく、分子生物学や放射線等の専門家とコミュニケーションを図りながら、生物学・医学的側面と統計学的側面からの議論することが重要である。

引用文献

- 1) 佐久間昭, 薬の効果・逆効果—臨床薬理学入門, 講談社 (1981)
- 2) 杉田克生 (編集), 学校での放射線リスク教育ガイドブック, 科学研究費補助金基盤研究 (B)「レギュラトリーサイエンスを導入した放射線教育プログラム開発ならびに教員養成」報告書 (2017)
- 3) 吉村功, 医学・薬学・健康の統計学—理論の実用に向けて, サイエンティスト社 (2009)
- 4) 矢野栄二 (翻訳), ロスマンの疫学—科学的思考への誘い, 篠原出版新社 (2013)
- 5) 丹後俊郎, 統計モデル入門 (医学統計学シリーズ), 朝倉書店 (2000)
- 6) 柴田義貞, 放射線リスク, 行動計量学, **43**, 35-43 (2016)
- 7) International Commission on Radiological Protection (ICRP), Recommendations of the International Commission on Radiological Protection, *Ann ICRP*, **21**, 1-3 (1991)
- 8) ICRP, The 2007 Recommendations of the international commission on radiological protection (ICRP Publication 103), *Ann ICRP*, **37** (2-4) 1-332 (2007)
- 9) Box GEP, Science and Statistics, *J Am Stat Assoc.*, **71**, 791-799 (1976)