

新連載 養護教諭のための ヘルスリテラシー講座
第2回 信頼できる情報としてのエビデンス

聖路加看護大学 保健医療社会学・看護情報学 教授 中山 和弘

信頼できるとは誤差の少ないこと

前回、ヘルスリテラシーとは、情報に基づいて意思決定し、自分の健康のために一番適した行動に移せる力であり、「健康を決める力」であると紹介しました。そのためには、まず、膨大なデータの中から信頼できる情報を手に入れる力が必要です。

信頼できない情報とは、例えば「好きなだけ食べて 10kg やせる」というような、被害やトラブルに遭う恐れがあるものです。では、そのような恐れのない信頼とはどのようなものでしょうか。みなさんにとって、信頼できると思う人はどのような人か考えてみてください。例えば、言うことが相手によってコロコロ変わったり、判断がぶれたりする人はどうでしょう。信頼できる人とは、言動や行動が一貫している人ではないでしょうか。多少考えが偏った発言をする人でも、いつもそうなら、それはそれで信頼できる人ともいえません。人に限らず、時計や体重計などのように、一般的に物事の信頼性は、一貫性すなわち誤差が少ないことで表されます。私たちが観察したり測定したりすることには、常に誤差がつきものです。したがって、「観測値」は次のように表されます。

$$\text{観測値} = \text{真の値} + \text{誤差}$$

観測値に占める誤差の割合が少ない場合に信頼性が高いといえます。

誤差について統計的に検討したものがエビデンス

多くの研究では、この真の値を知ろうとしています。例えば、ある調査で、運動部に所属した生徒は、そうでない生徒より将来的に喫煙者に 1.5 倍なりにくいというような数値が計算されたとします。その時、この値がどれだけ信頼できるかが問題です。真の値は1なのに、研究の対象や方法によっては、誤差が加わって 1.5 になることがあります。そのように発生してしまう誤差には2種類あります。まったく偶然のもの（偶然誤差）と何らかの理由があるもの（系統誤差）です。前者は仕方ありませんが、後者は研究方法が原因となっていることが多く、バイアス（偏り）と呼ばれます。その誤差の発生の仕方を知り、それを取り除くなどのコントロールをするのが統計学や疫学です。1.5 という数値がどれだけ信頼できるのかについては、「95%信頼区間」がよく使われます。

例えば、それが 1.4 ~ 1.6 という幅になったとします。それは、別の人々を対象に何度も同じ調査を繰り返しても、真の値は 95%の確率でその幅に入っているだろうという意味

です。そうであれば、1から十分に離れた値だということと（これは統計的検定で有意だという意味と同じです）、1.5がそれほど大きな誤差を持っていないことが確認できます。このようにして統計的な検討を行って生み出された信頼できる研究結果をエビデンスと呼びます。それは、証拠という意味の英語であり、科学的根拠という意味で使われます。

真の値をゆがめてしまうバイアスに注意

例えば、テレビ番組で、ある食品のダイエット効果を調べるために5人を対象にした実験を行って、平均体重が1か月で4kg減ったとします。その数値にどれだけバイアスがかかっているかを考えてみましょう。

まず、対象者を選ぶ時です。その5人はどのような人々を代表して実験に参加しているのでしょうか。アルバイトに募集した人でしょうかし、結果が出やすいようにお正月などで一時的に太った人が選ばれていて、放っておいても元に戻るのかもしれませんが（選択バイアス）。体重計の設定が正確でないとか、テレビカメラを操作して縦長に映しているかもしれません（測定バイアス）。期待に応じてやせなくてはいけないというプレッシャーが原因で、生活全般に変化が起きていたことも考えられます（交絡バイアス）。

このようにバイアスには主に3つあるのですが、特に交絡（混乱・混同という意味）は、やせた原因の真犯人はプレッシャーなのに、ニセの犯人をでっち上げるようなものですから要注意です。同様に、よく知られているのは、薬の治療におけるプラセボ（偽薬）効果で、これはどんな治療でも効くと思えば効くという心理的な効果が働くことです。

バイアスの少なさを意味するエビデンスレベル

エビデンスでは、これらのバイアスが排除できているレベルで信頼性が変わってくるので、それをエビデンスレベルといいます（図）。

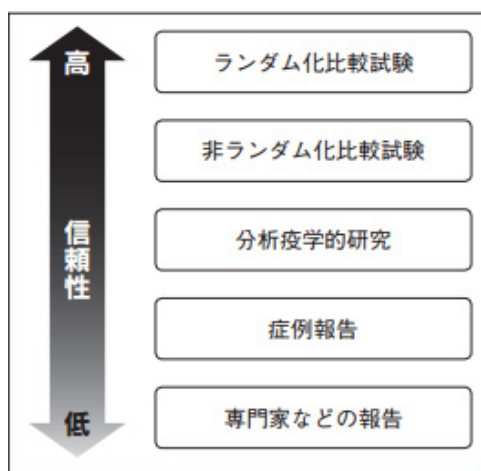


図 エビデンスレベルで変わる信頼性

低い方から順に見ていくと、まず専門家などのデータに基づかない意見や報告です。データが示されなければ、証拠にならないということです。次が、数の少ない症例報告で、これは検証すべき仮説を提案する目的では大切ですが、統計的に検証されたとはいえません。

きちんとデータを統計的に分析して検証した研究は、分析疫学的研究からになります。この研究の場合は、まったく観察するだけで、対象者に対して何か介入することはありません。2種類の方法がありますが、1つの方法は、原因を知るために過去にさかのぼるもので、例えば現在、肥満の人とやせの人に、原因と考えられる過去の状況（例：運動や食事）を聞いて違いがないかを比較するものです。もう1つの方法は、原因が未来に影響を与えるかを見るもので、対象者の現在の肥満の状況を聞いておいて、それに変化を与えると予想される状況（例：運動や食事）を定期的に調べて、違いが出てくるかを比較するものです。

さらにレベルが高いのは、比較試験です。研究対象者に介入して、すなわち原因を作ってもらったり（例：ある食品をとる）、なくしてもらったり（例：ある食品をとらない）して、差ができるか（例：肥満度）を実験する方法です。介入の有り無しの2つのグループを作りますが、その時、くじ引きのような方法で分けるランダム化比較試験の方が優れています。そうでない非ランダム化比較試験の場合は、グループ分けに研究者の意図が入り込みバイアスが発生する可能性があるからです。また、プラセボ効果を取り除く方法として、本物の食品とニセ物の食品（プラセボ）の2グループに分ける時、対象者も研究者もどちらの食品を食べているのかを知らせないということが行われています（マスキングといえます）。

また、観測値は、対象者の数が少ないほど、偶然にでも差が大きく出てしまう可能性があります。例えば、日本人女性からランダムに5人ずつ選んで平均体重を比較すると、4 kg以上の差ができる確率は1/2ほどもあります。この確率は20人ずつだと1/4、40人ずつだと3/100と、人数が多いほど小さくなります。したがって、人数が多いほどエビデンスの信頼性を高めることができます。

更に詳しくお知りになりたい方は、「健康を決める力」(<http://www.healthliteracy.jp/>)の「信頼できる情報とは何か」をご覧ください。