

## 最終講義

## 数理で考える疾病予防

—無駄な医療をなくしたい—

岡田正彦

元 新潟大学大学院医歯学総合研究科

予防医療学分野

**Mathematical Studies on the Preventive Medicine: My Effort to  
Decrease Overdiagnoses and Overtreatments**

Masahiko OKADA

*Former Professor of the Division of Preventive Medicine,  
Graduate School of Medical and Dental Sciences,  
Niigata University*

## 要 旨

予防医療学の基礎作りには、患者データの集積と、その数理解析が欠かせない。本稿では、私が取り組んできた数理による疾病予防への戦略について述べる。また、諸外国で行なわれた数々の大規模調査から明らかにされた過剰な医療の実態をまとめ、これからの医療がいかにあるべきかについても考察してみたい。私がまず取り組んだ研究は、有用なデータと無用なデータを数学的に区別する方法についてであった。診療に役立つデータ項目の組合せは、以下のような条件を同時に満たすものでなければならない。①少数の疾患で特異的に陽性になること、②他のデータと情報が重複していないこと。私が見出した公式「相関重みつきエントロピー法」は、両条件を同時に満たすことができる。これらの努力と並行して、患者情報をコンピューターに集積する技術についても研究を行ってきた。欧米においても、この分野の研究は盛んに行なわれ、その成果は現在の電子カルテにつながっている。ところが、しばらくして、まったく新しい発想で大規模臨床試験や大規模コホート調査が行なわれるようになり、驚くべき結果が相次いで発表された。分かったことをひと言でまとめれば、現代医療に総死亡を減らす効果は認められないこと、および無目的に患者データをコンピューターで集めても何ら役に立つことはないという厳粛な事実であった。医療のあり方を根本から問い直すべき時に来ているように思われる。

キーワード：予防医療学，数理，ビッグデータ，大規模調査，総死亡

Reprint requests to: Masahiko OKADA  
Mizuno Memorial Hospital  
6-24-13 Nishiarai Adachi-ku,  
Tokyo 123-0841 Japan

別刷請求先：  
〒123-0841 東京都足立区西新井6-24-13  
水野記念病院・水野介護老人保健施設 岡田正彦

## 緒 言

「医療情報が氾濫している」と言われるようになって、すでに久しい。確かに、整理されないままの情報は、その量がいくら多くとも役に立つことは少ない。それどころか、類似した情報ばかりが多くなってくると、結論を誤る可能性が逆に高まることも数学的に示される。私が学生だったころ、「同じ目的の検査をできるだけ多く行うように。なぜなら1つの検査に不備があっても、別のデータで診断ができるから・・・」という講義を受けた記憶がある。このような発想は明らかに間違っている。

私の最終講義では、数理による疾病予防の戦略について述べ、次いで、諸外国で行なわれた数多くの大規模調査の数理的背景と、過剰な医療の実態を明らかにし、真の医療がいかにあるべきかを考察してみたい。

### 重複した情報の整理

情報の根源的な数量を数量化することができる重要な公式がある。「根源的」とは、これ以上、数字を小さくできないという意味だが、これをエントロピーと呼び、単位をビットで表す。ある検査Xを考え、それが陽性になる確率を $p$ 、陰性になる確率を $q$ とする（ただし $p+q=1$ とする）。もし疾患Aに対して常に $p=1$ であれば、 $q=0$ となる。この検査は、Aを診断する上で理想的といえるだろう（ほかの疾患は考えないとして）。このとき、エントロピー $H$ は、

$$H = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

と計算される。 $p=0$ 、 $q=1$ となる場合も同様である。一方、 $p=0.5$ 、 $q=0.5$ では、

$$H = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

となる。このとき、 $H$ は最大値をとる。図1は、こうして求めた（あらゆる） $p$ と $H$ の関係をグラフ

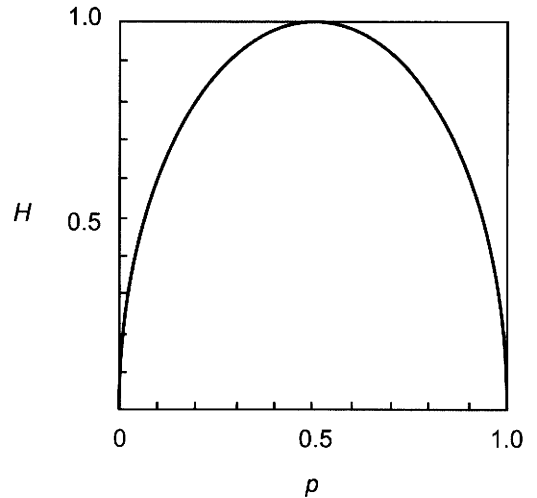


図1 出現確率 $p$ とエントロピー $H$ の関係

にしたものである。対象疾患が2つ以上の場合も同様に数量化されるが、式が複雑になるため、詳細は省略する。

・・・と、ここまでは情報理論の教科書に載っている基本的な説明だが、実際の臨床データは、種類が多く、かつ互いに情報の重複がある（相関が高い）。そこで私は、診療に役立つデータ項目の組合せは、以下のような条件を同時に満たすものでなければならないはずと考えた。

- ①少数の疾患で特異的に陽性になること
- ②他のデータと情報が重複しないこと

最終的に、両者を同時に満たす新たな公式「相関重みつきエントロピー法」を見出すことができたが、そのきっかけは、ある日、突然、この2つの条件を同時に満たす3D画像が脳裏に浮かんだことに始まる。その絵をすぐにメモ用紙に書きとめ、数式に置き換えては実際のデータで試してみる、という作業を繰り返していったのである。これを論文にして米国の専門誌に投稿したところ、欠点がない理論と評され、書き直しを要求されることなく採択となった<sup>1)</sup>。投稿した論文がそのまま印刷されたのは、私の研究者人生において、これが最初で最後であった。のちに、これがきっかけと

なり、同論文が掲載された学会誌の副編集長を長く務めることになる。

### ビッグデータの処理技術

最近、ビッグデータという言葉が流行っている。コンピューターの記憶容量と処理能力の飛躍的な向上により、テラ（1兆）バイト、ペタ（1000兆）バイト、エクサ（100京）バイトを超える巨大データが集積されている状況を指す。生物医学の分野でも、巨大なデータを網羅的に検索することで、何か新たな発見が生まれるのではないかと、という期待が高まっている。

30年ほど前、医療の分野でもビッグデータの時代が来ると感じた私は、氾濫する医療情報に対処するため、無駄なデータを排除するだけでなく、個々の数値、文字、画像なども圧縮して保存する必要があるのではないかと考えた。たとえば、256という値をコンピューターに入力する場合を考えてみよう。3桁の10進数から成っているが、コンピューター内には1と0の数字（2進数）しか存在しないため、

2 → 0000 0010

5 → 0000 0101

6 → 0000 0110

と記録される（このコード体系はASCIIと呼ばれる）。つまりコンピューターでは、256という情報を表現するために、 $8 \times 3 = 24$ 個もの桁数の2進数が使われている。2進数の1桁を1ビットというが、2進数が使われているのは、これが情報の最小単位であり、雑音などの影響を受けにくいからである。上位4桁はアルファベットや記号も一緒にコード化するために必要なものである。0から9までの数字を区別するだけであれば、下位4桁で済むが、それでも必要な情報量は計12ビットになる。

その昔、Huffmanという数学者が、情報を最少のビット数で記録するための理論構築を行った。尿蛋白の検査データを例に具体的に考えてみよう。たとえば結果が{-, ±, +1, +2, +3, +4}の6通りあり、大部分が-で、以下、順に出現

頻度が低くなり、+4は稀にしか現れないとする。その確率が、

{0.20, 0.19, 0.18, 0.17, 0.15, 0.10}

だったとする。これらは、ASCIIコードのように同じ長さ（固定長）のコードに置き換えるよりも、頻度の高いデータほど短い（ビット数の少ない）コードを割り当てたほうが、トータルの容量は少なくなる。このようなデータ系列は、Huffmanの考えたアルゴリズム（数学的な処理手順）によれば、

{10, 11, 000, 001, 020, 0110}

という2進数をそれぞれ変換コードとして割り当てて記録（あるいは通信）したときに最少となる。これをハフマン・コードと呼び、各コードのビット数が一定でないが、互いに重複しないように生成されるため、元のデータ系列が正確に復元できる。ただしデータ系列に固有なため、常にコード表が手元にないと復元はできない。

医療情報に適した新たなデータ圧縮技術の研究に取り組んだ私は、実験対象としてCT画像を選んだ。まず実験で分かったのは、ハフマン・コードは、コード表自体のサイズが予想外に大きいこと、および変換に要する時間も無視できないことであった。理論的に究極の圧縮ができることに間違いはないが、元のデータを復元するための効率を考えると、必ずしも最善な方法ではなかったのである。そこで考えた方法が、以下に示すオリジナルのパケット・コード法であった。

CTなどの画像データの場合、隣り合った点（ピクセル）は互いに似通った値を取ることが多い。そこで、まず各ピクセルの値を画像の片隅から順に、全体について引き算していく。すると差の多くはゼロ、またはそれに近い値になる。差が大きくなるのは、骨と脂肪の境目などにおいてである。この数値は、先に例示した尿検査の結果に似ている。

これを2進数に置き換え、かつ2ビットずつ（パケット）に切り分けていく。先頭パケットの1ビット目は常に1とし（ヘッダー）、2つ目以降のパケットでは、1ビット目をゼロ（リンカー）としておく。実際のCT画像から得た引き算の結果

と、その出現確率、および変換されたパケット・コードを以下に示す。

0	0.38705	10
+1	0.03039	11
-1	0.03006	1000
+2	0.02920	1100
-2	0.02895	1001
+3	0.02723	1101
-3	0.02704	100000
+4	0.02518	110000
-4	0.02512	100100

(以下、省略)

ゴシック文字はヘッダーまたはリンカーである。各パケットは2ビットで固定されており、かつヘッダー(1)とリンカー(0)により元のデータが容易に(変換コード表なしに)識別できるのである。

以下、同様に3ビット長、4ビット長のパケットについても検討を行なったが、実際のデータで圧縮率を求めたところ、2ビット長パケットがもっとも優れていることが分かった。表1に、パケットの長さによる圧縮率の違いをハフマン・コードのそれと共に提示した。この方法により、CT画像が5分の1にも圧縮でき、かつ完全な原画像を瞬時に再生することができるようになった<sup>2)</sup>。

### 情報検索の最短路

データ量が極端に増大していくと、必要とする情報を検索するための時間も大きな問題となる。現代のインターネットでは当たり前になっている検索技術も、まだ萌芽的な状態だった時代のことである。

通常、情報をデータベースに記録する場合、情報本体とその識別子(キーワード)を別ファイルにして保存する。たとえば検査結果(情報本体)と患者ID(キーワード)をそれぞれイメージすると分かりやすい。情報検索技術の1つに、キーワードを最速で探し出すためのB木という手法がある。しかし、この方法では、重複したキーワードを処理することができない。そこで、同法を改

良し、キーワードが2個以上あっても検索ができる新しいアルゴリズムを開発した。

本法の検索能力を検証するため、73,200人分のデータからなる集団検診データベースを対象に処理を行った。この検診は不特定多数の人を対象にしていたことから、IDが付与されていなかった。そこで、名前(ファーストネーム)、誕生日、それに電話番号から成る20バイトのキーワードを構成し、同一人が5年間に何回、検診を受けているかを調べてみた。その結果、実受診者数が39,421人おり、そのうち1回だけ受診した者21,542人、2回受診した者8,255人、3回受診した者4,933人、4回受診した者3,106人、5年連続して受診した者1,585人であることが分かった。処理時間を無視すれば、このような集計を行なう方法はいろいろある。しかし、本アルゴリズムでは、ほぼ最短で検索ができるのである。

この方法を報じた論文は、米国の専門誌に掲載されたが<sup>3)</sup>、現代のビッグデータ処理技術に多少の貢献を果たしたものと確信している。

一連の研究が成立しえた背景の1つは、当時、医療へのコンピューター応用が始まっていたことである。もちろん現代のようなインターネットはなく、パソコンもない時代であった。超大型コンピューター、およびミニコンピューターと呼ばれる工業用・研究用の装置のいずれかしか存在せず、前者は大きな金融機関や航空機の予約システムなどに限られ、医療への応用は、もっぱら後者によるものであった。

私自身、日本で2台目とされる米国製の最新鋭ミニコンピューター(PDP-12)を使える機会をえて、患者情報を集積する技術の研究を並行して行っていた<sup>4)~6)</sup>。患者データを手当たりしだいコンピューターに記録させておけば、いつかは医療の進歩に役立つはず、という期待があったからである。海外にも同じ発想を持った研究者は多く、大いに盛り上がっていた。それらの努力が、現代の電子カルテ技術の基礎作りに貢献したと言えるだろう。

表1 パケット・コードとハフマン・コードの比較

	ビット長
ハフマン・コード	4.791
2:1 パケット・コード	5.266
3:1 パケット・コード	5.418
4:1 パケット・コード	5.988

### 大規模調査の衝撃

1990年、検査診断学講座を担当することになったが、この頃より世界では大規模臨床試験、大規模コホート調査が続々と行なわれ、驚くべき結果が相次いで発表されるようになっていた。その先駆けは、糖尿病治療薬のスルフォニル尿素剤についての調査報告であった<sup>7)8)</sup>。一連の調査は1961年から1975年にかけて行なわれ、1970年代に何報かの論文が発表された。以下は、その概要である。

まず823人のボランティアが4群（ダイエットだけの群：205人、固定量のインスリンを注射する群：210人、血糖値に合わせてインスリン注射量を加減する群：204人、スルフォニル尿素剤服用群：204）に無作為に分けられた。各群を9年以上に亘って追跡し、心血管障害による死亡と総死亡をそれぞれ調べたのである。その結果、心血管疾患による死亡はダイエット群で10例、固定インスリン群で13例、加減インスリン群で12例、スルフォニル尿素剤で26例となった。同じく総死亡は、それぞれ21、20、18、30例だった。イン

スリンは使っても使わなくとも心血管疾患による死亡率、および総死亡に差はなかったのである。スルフォニル尿素剤に至っては、ダイエット群に比べて圧倒的に死亡率が高くなってしまっていた。

一連の論文が発表された当時、なぜか日本ではあまり話題にならなかったが、私にとって2つの意味で大きな衝撃だった。1つは、従来から教科書にも記載されていた医学の常識が、一遍の論文で脆くも崩れ去ってしまったことで、もう1つは患者データを手当りしだいコンピューターに記録させておけば、いつか医療の進歩に役立つのではないか、という淡い期待が脆くも崩れ去ったことであった。

この研究が先駆けとなり、その後、医療のあらゆる分野で大規模調査が行なわれ、膨大なエビデンスが提示されてきたのは、ご存知の通りである。多くの調査で、教科書的な常識が覆される結果となったのであるが、本文の巻末に、その概要をまとめた拙著を参考文献として挙げたので、ご参照いただければ幸いである<sup>9) - 11)</sup>。

## 結 語

目的もなく網羅的に収集しただけの医療データは、結局、何の役に立つこともなかった。電子カルテの時代となり、再び同じ失敗を繰り返さないでほしいものである。

一方、明確な目的の下に行なわれた数々の大規模調査には、医療の常識を覆す大きなインパクトがある。結論をひと言でまとめれば、現代医療には総死亡を減ずるほどの効果は認められないということなのである。この厳粛な事実を真摯に受け止め、今、医療のあり方を根本から考え直すべきなのではないだろうか。

## 謝 辞

本文で述べた研究には多くの方々の協力が不可欠であった。心からの謝意を捧げるが、あまりにも人数が多く、個々の氏名は省略させていただく。

## 文 献

- 1) Okada M: A method for clinical data reduction based on "weighted entropy". IEEE Trans Biomed Eng 25: 462 - 467, 1978.
- 2) Okada M: Storage volume of computer tomography images can be reduced by a factor of five. Comput Meth Prog Biomed 20: 151 - 157, 1985.
- 3) Okada Ma and Okada Mi: An algorithm to count the number of repeated patient data entries with B tree. Comput. Biomed Res 18: 121 - 124, 1985.
- 4) Okada M, Maruyama N, Kanda T, Shirakawa K and Katagiri T: Medical data base system with an ability of automated diagnosis. Comput Prog Biomed 7: 163 - 170, 1977.
- 5) Okada Ma and Okada Mi: An algorithm to manage variable-length records for highly portable clinical data base systems. Comput Biomed Res 19: 224 - 236, 1986.
- 6) Okada Ma and Okada Mi: An evaluation of data items for the filing system of a gastric cancer screening project. Med Inform 11: 177 - 183, 1986.
- 7) Meinert CL, Knatterud GL, Prout TE and Klimt CR: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes, II. Mortality results, Diabetes 19: 789 - 830, 1970.
- 8) Miranda VP, Rodrigues MJ, Goncalves FR and Nunes JP: Effects of hypoglycemic agents on mortality and major cardiovascular outcomes in patients with type 2 diabetes mellitus: a narrative review. Rev Port Cardiol 28: 1099 - 1119, 2009.
- 9) 岡田正彦：がん検診の大罪。新潮選書, 2008.
- 10) 岡田正彦：薬なしで生きる。技術評論社, 2009.
- 11) 岡田正彦：検診で寿命は延びない。PHP新書, 2010.