

## Longitudinal Data の統計解析上の留意点

本田 純 久

長崎大学大学院医歯薬学総合研究科附属原爆後障害医療研究施設放射線疫学研究分野

同一の対象者を長期にわたり追跡し、対象者の曝露状況や疾病状況、交絡因子の状況の変化を調べる longitudinal study (縦断研究) では、データの構造を正しく把握し、正しい統計解析を行う必要がある。本稿では、縦断研究において、個人の対応を無視してデータ解析を行ったために生じるバイアスについて述べ、データ構造を正しく把握することの重要性について説明する。

### 縦断データの構造

図 1 は仮想的な数値例について、12人の対象者の年齢と収縮期血圧の値を散布図に描いたものである。グラフからも明らかなように、年齢が高くなるにつれて血圧値が高くなる傾向がみられる。ピアソンの相関係数の値は0.80 ( $p < 0.01$ ) となり、統計解析の結果からも年齢と血圧値の間に強い正の相関がみられる。

ところが、同じデータが実は食生活や運動などの生活習慣の改善による血圧値の変化をみるために、6人の対象者を1年間フォローアップした縦断研究から得られたデータであったとしたらどうなるであろう。図 2 は各対象者の血圧値の変化を直線で結んだものである。対象者の血圧値はフォローアップ期間中に平均で8.2mmHg 低下している。また対応のある t 検定 (paired t-test) の結果から、統計的に有意に ( $p = 0.01$ ) 血圧値が低下していることがわかる。

さらに、同じデータが実は2年の間隔で行われた住民検診の結果であったとしたらどうなるであろう。図 3 は各対象者の2年間の血圧値の変化を直線で結んだものである。対象者の血圧値は2年間に平均で8.8mmHg 上昇しており、対応のある t 検定を行った結果でも、統計的に有意に ( $p < 0.01$ ) 血圧値が上昇していることがわかる。同じ12組の

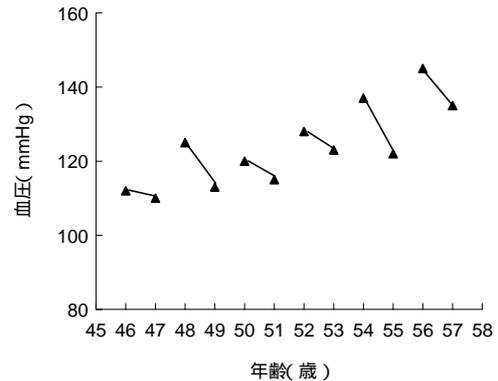


図 2 年齢と血圧値に関する縦断的データの例 1

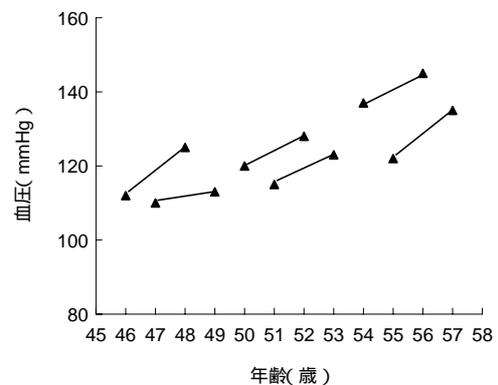


図 3 年齢と血圧値に関する縦断的データの例 2

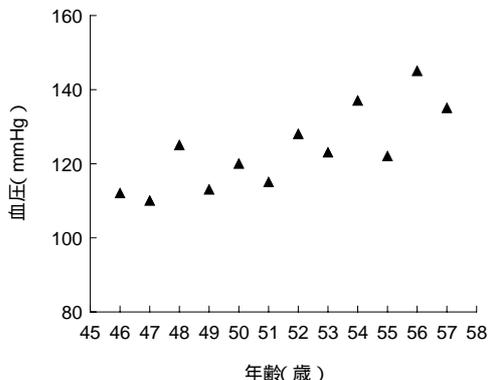


図 1 年齢と血圧値に関する仮想的データ

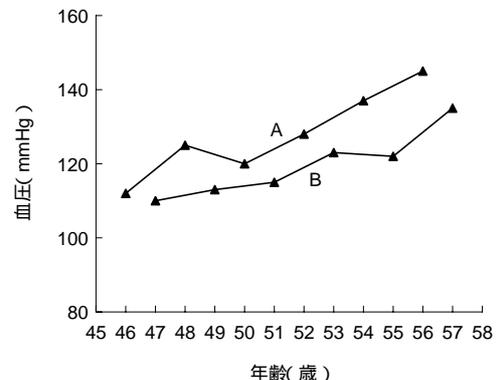


図 4 年齢と血圧値に関する縦断的データの例 3

年齢と血圧値のデータであっても、データの構造が異なると反対の結果になることを図2と図3の例は示している。また正しい統計解析の方法を選ぶことも重要である。図2と図3の例について、個人の対応を無視した通常のt検定を行うとp値はそれぞれ0.21と0.17となり、統計的に有意な差とはならない。縦断データではデータの構造を理解して正しい統計解析を行うことが重要である。

もし図1のデータが2人の対象者を2年の間隔で6回の検診を行った結果であったとしたらどうなるであろう。図4は2人の対象者A、Bの10年間における血圧値の変化を直線で結んだものである。各対象者の年齢と血圧値について単回帰分析を行うと、対象者Aは1年あたり3.0mmHg血圧値が上昇し、対象者Bは2.3mmHg血圧値が上昇していることがわかる。このように同一対象者を繰り返し測定したデータのことをlongitudinal data(縦断データ)とかrepeated measurement dataという。

#### 疫学研究における縦断データ

疫学研究の代表的な研究デザインであるコホート研究は、対象者を長期にわたりフォローアップすることから縦断研究として捉えることができよう。しかしこれまでのコホート研究では、縦断的にデータを集めるのは新たな疾病の発生という疾病状況の変化のみであることが多く、フォローアップ期間中の曝露因子や交絡因子の状況の変化はあまり考慮されなかった。統計解析においても、多くの場合は研究開始時点での曝露因子や交絡因子の状況をもとに解析が行われていた。実際には、疫学研究で扱う多くの曝露因子や交絡因子(例えば、喫煙、飲酒、食事、運動などの生活

習慣、血圧値、コレステロール値、体重などの身体状況、ストレスや抑うつなどの精神的・心理的状況、大気汚染や水質汚染などの環境疫学における曝露状況、産業保健における職業曝露の状況)は、時間の経過にともない変化する(Rothman and Greenland, 1998)。変化する曝露因子や交絡因子の状況をもとに、疾病発生に与える曝露の影響の大きさ(疾病発生率比やリスク比)を推定するための新たな統計手法(例えば、Generalized Estimating Equationsや時間依存性Cox回帰モデル、混合効果モデル、G-estimation)の開発が疫学研究の分野でも進んでいる(Robins et al., 1992; Verbeke and Molenberghs, 2000; Diggle et al., 2002)。またSASやSPSSなどの統計解析ソフトでも、これらの統計解析を行うことが可能となった。今後疫学研究における縦断データの重要性はますます認識され、統計手法の応用も進むであろう。

#### 文 献

- Diggle PJ, Heagerty P, Liang KY, Zeger SL: Analysis of Longitudinal Data, 2nd edition. Oxford University Press, New York, 2002.
- Robins JM, Blevins D, Ritter G, Wulfsohn M: G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. *Epidemiology* 3: 1992, 319-336.
- Rothman KJ, Greenland S: Modern Epidemiology, 2nd edition. Lippincott-Raven, Philadelphia, 1998.
- Verbeke G, Molenberghs G: Linear Mixed Models for Longitudinal Data. Springer-Verlag, New York, 2000.