The **THAI**
Journal of **SURGERY**
Official Publication of the Royal College of Surgeons of Thailand

Vol. 31 January - March 2010 No. 1

Review Article

Variability and Randomness in Medical Research

Panuwat Lertsithichai, MD, MSc. (Medical Statistics)
Department of Surgery, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand.

Abstract **In the present article, the first in a series of articles on the principles of statistics for surgeons, the ideas of variability and randomness in medical observations are developed. Variability can be partitioned into two components: systematic and random. The detection of systematic variability is the objective of scientific investigations, while random variability can be defined as the residual variability after systematic variability has been taken into account. That is, if the residual variability can be shown to satisfy, approximately, the statistical criteria of randomness. In the present article, we provide two examples in which the residual variation is shown graphically to behave like random variation. In practical applications, random variation should not be defined as having "no cause". Similarly, randomness should not be defined as complete unpredictability. In later articles in the series, we will explore how the ideas of variability are used in medical research and statistical analyses.**

Keywords: **Predictability, random variability, statistics, systematic variability**

INTRODUCTION

Variability is the norm in everyday observations, including medical observations on patients. For example, patients undergoing a subtotal gastrectomy can have a variety of outcomes, such as death, survival with anastomotic leakage, or survival with full recovery. Surgical patients can be men, women, young, old, with or without concurrent illnesses, rich, poor, of European or Asian ancestry, and so on. Variability is so ubiquitous that it would be remarkable if any two patients are found to have the same response to a treatment, in all detail, even if they are identical twins. It is to be expected, therefore, that variability will play an important role in the process of scientific inference. In fact, the appropriate analysis of the sources of variation is the key to important medical discoveries.[1]

Correspondence address : *Panuwat Lertsithichai, MD, Department of Surgery, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand. Telephone: 066 2 2011315; Fax: 066 2 2011316; E-mail: raplt@mahidol.ac.th*

Sources of variability[2]

Medical observations on patients are variable due to a number of sources. Variability can be systematic, i.e. consistently associated with some measurable "factors". For example, the observation "height", or the measure of height, in the mature adult can vary by such factors as gender (males are often taller than females in a given population), age (we tend to be somewhat shorter with age), and socioeconomic status (the rich may be taller than the poor). However, factors associated with or related to height may or may not be the "cause" of tallness or shortness. Age is not the immediate cause of the variation in height; the bony changes in the elderly, for example the increase in curvature and the collapse of the spinal column, are more directly responsible for the decrease in height. But the important idea is that systematic variability can be predicted to a certain extent. This "predictability" applies even to the individual patient.[3] The objective of medical research is often to determine the systematic component, especially the cause, of the variability in outcome.

Variability can be causal. (It is beyond the scope of the present article to argue for the "reality" - in the philosophical or epistemological sense - of causality.[4] We will simply accept causality as real in the common, every day understanding of the term). In a sense, medical research strives for causal explanation of medical phenomena. That a causal explanation can rarely be established with complete confidence is not a present concern, however. Often, a systematic source of variation is a marker of an underlying cause, and occasionally subsequent studies may establish the identity between the systematic "factor" and the "cause". For example, the identification of the bacteria *Helicobacter pylori*, initially seen as a colonizing organism in peptic ulcers, as a "cause" of peptic ulcers is well known.[5] Mutations in the p53 gene can only be seen as marker of increased risk ("risk factor") for some cancers, until subsequent studies of the function of the p53 protein in the cellular apoptosis cascade and cell cycle arrest established the p53 gene mutation as an important "cause" of cancer.[6] The randomized controlled trial (RCT) can be viewed as an attempt to experimentally verify suspected causal sources of variability in outcome. That is, differences, or variability, in the outcomes between two or more treatment groups in an RCT can be causally attributed to the differences, or variability, in the treatment factor.

Variability can be considered random. Randomness can be defined in a mathematical manner,[7] without reference to the underlying "causes" of "random" variation (see later). Some authors equate randomness, by definition, with "no cause".[8] This view is surely a mistake. An example of randomness within a causal framework is the random motion of a gas molecule due to collisions with other molecules, as modeled in classical statistical mechanics.[9] Although the laws of physics in this case are causal, the result for the entire system is random motion. In an analogous manner, variation in medical observations can be regarded as random if such variation approximately satisfies some mathematical properties. But this by no means implies that such variation is causeless. In fact, a multitude of causes, many of which we choose to ignore at any given time, might combine to produce random variation, much as the random motion of a molecule is caused by a countless number of molecular collisions. In practice, or at least in statistical practice, any variation not accounted for by the systematic factors under study - the "residual variation" - in the outcome of any particular research study or experiment can be regarded as random under some suitable statistical model, even if we suspect that some unexamined causes underlie the residual variation. Importantly, we do not deny that certain components of the random variation may be causeless, but we entertain the possibility that a large part of the so-called random variation could be due to unknown causal factors.

In a world where no "real" random variation exists, but only systematic variation, perfect predictability could be achieved in principle. (The words "prediction" and "estimation" will be used almost interchangeably in the present article, even though they do not have the same meaning in conventional statistics). This Laplacian view of the world (named after the famous French mathematician Pierre-Simon Laplace, 1749-1827, whose clear and eloquently expressed idea to that effect was often quoted)[10] was a dominant one in the late 18th and 19th century Europe. For example, if the result of an operation can only vary by age and gender, then the information that the patient is male and 68 years old is all that is required to provide a completely accurate prediction of the outcome for that patient. Statistics as a discipline is

then a set of semi-empirical methods for extracting useful but imperfect information, to be superseded when scientists finally discover the full set of systematic factors underlying all the variation in any given set of observations. In the present era, most scientists no longer believe in such a simplistic view of the world, and so the surgeon cannot avoid but to take the idea of random variation seriously.

Examples of random variation in statistical models

The idea of random variation is so important and fundamental to statistics and the statistical view of the world that some sophisticated examples must be given. Consider once again the observation of height, but in a group of subjects aged between 10 and 20 years. Suppose that we are interested in the systematic variation of height with age. That is, "age" is the systematic factor of interest for the observation "height" in the present example. There might be other systematic factors associated with the variation in height, such as gender and socioeconomic status mentioned earlier, as well as other factors known or unknown, but we choose to ignore them. After the variation in height due to age has been taken into account, any residual variation left over might be considered random variation, if under some suitable statistical model it can be so interpreted.

Quantitatively, and specifically, let us consider the following model of the observed measurement, height:

Height (cm.)
= [Constant × Age (years)] + Residual.

That is, the variation in height is linearly related to the variation in age (for the age group 10 to 20 years) and a residual variation. That there must be some residual variation is obvious, since no matter how perfect the measurements of height and age, age alone cannot perfectly determine height. Such a model equation specifies one part of the statistical model. Other model assumptions include the distribution model of the height, and the sampling model in a population from which the sample of observations is drawn. If we assume that each observation is randomly and independently drawn (i.e. random sampling) from a defined population - such as the Thai population - and that the variable "height" has a "Normal distribution" (i.e., a Bell-shaped distribution) in that population, then the residual variation in the sample can be considered random if its distribution is compatible with that of a zero-mean normal variate.

More clarification is probably needed. In pictures, if the measurement "height" in a random sample of subjects has, approximately, a Normal distribution as in Figure 1, and if the residual also has an approximate Normal distribution with mean zero as in Figure 2, then the residual variation can be considered random, regardless of the underlying causes of the residual variation. More generally, if a graph of the residual plotted against age is symmetric about the value zero (Figure 3), then the residual variation might be approximately random. Such graphical checks on the model for the data can be done using any statistical software package for personal computers, if the sample size is large enough.

Another example may help to further the understanding of the statistical idea of random variation. In a yet unpublished retrospective cohort study of 1,852 inguinal hernia repairs in 1,533 patients, it was found that 1,017 repairs were for the right side. In other words, the percentage of right-sided inguinal hernias in the study was 54.9% (1,017/1,852). The systematic factors of interest included age (perhaps the elderly has a propensity to right-sided hernias) and gender (perhaps males are more likely to have right-sided hernias). That is, we want a statistical model explaining the variation in the side of inguinal hernia occurrence with age and gender as systematic factors, along with a residual variation, which we may write schematically as:

Side of hernia occurrence
= f(Age + Gender) + Residual.

In the present example, we have generalized the contribution of the systematic factors in the form of a suitable function f(...), instead of inserting the factors unchanged directly into the equation, as was done for the outcome "height". This is because the outcome "side of hernia" is now a "binary" variable, with two "values": right or left, which we may code numerically as "1" and "0" for convenience. (In contrast, "height" is a continuous variable). Therefore, the statistical model for a binary outcome and the methods to "check" randomness will be somewhat different from those of height. In the following, we will again demonstrate a simple graphical criterion for checking whether the residual variation can be considered random under a suitable statistical model.

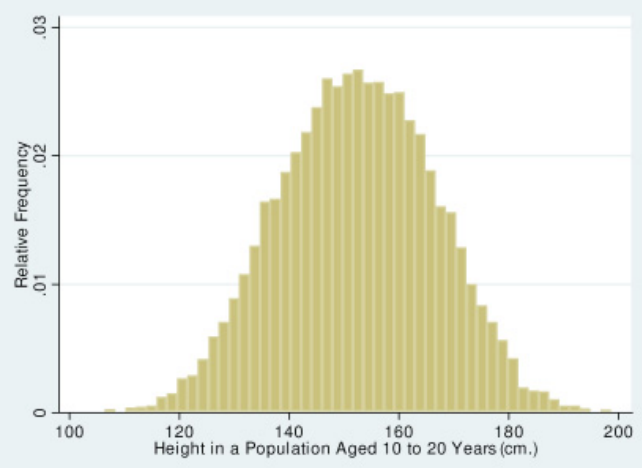

Figure 1 The histogram of a data set of 10,000 observations of height, randomly drawn from a simulated population with age uniformly distributed between 10 and 20 years. The variability of the observations is between 100 cm. to 200 cm.: a range of 100 cm.

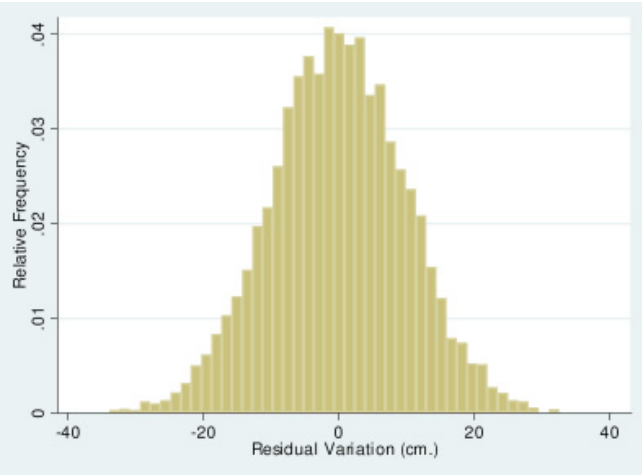

Figure 2 A histogram of the residuals (here labeled as "residual variation"), from the same data set as that of Figure 1, showing a symmetrical, bell-shaped distribution about the zero-value: the zero-mean Normal variate. Compared with Figure 1, the variability of the residuals is smaller (between - 35 cm. to + 35 cm.: a range of 70 cm.) because the variation due to age has been taken into account.

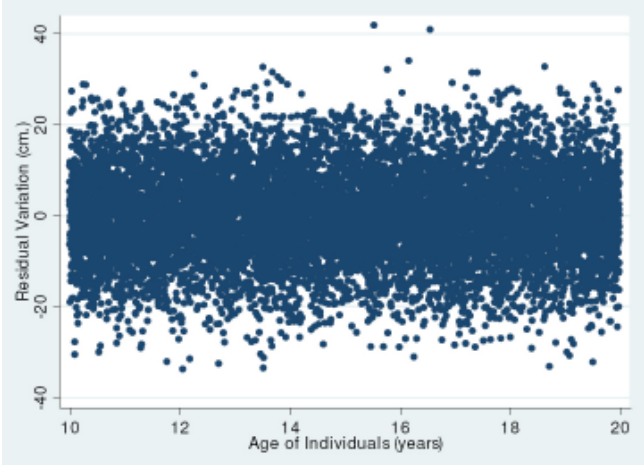

Figure 3 An alternative and more general graphical "check" of random residual variation: a "scatter" plot of residuals against the systematic factor of interest, "age". The data set displayed is the same as that in Figure 2. Note the symmetrical distribution about the value zero (in the horizontal direction) with no systematic pattern.

A common statistical modeling approach is to consider the variation in the side of each hernia to have arisen from a so-called "Bernoulli process" (named after the great Swiss mathematician, Jacob Bernoulli, 1654-1705, who wrote on the subject in his famous posthumous book, *Ars Conjectandi*).[11] This process is the same as that arising from the tossing of a coin. The "sidedness" of a given hernia is assumed independent of that of another, inasmuch as each tossing of the coin can be assumed similarly independent. (In fact, the sidedness of the hernia is only independent between patients. For example, a right-sided hernia in one patient should not influence the sidedness of the hernia in any other patient. However, the sidedness of hernias in a given patient may not be independent. If a patient develops a hernia on the right side, and subsequently had the hernia repaired, he or she may be more likely to develop another hernia on the left side in the future. However, for the sake of simplicity, we will assume that all hernia occurrences are "independent" in the study).

The combination of a sequence of Bernoulli processes, the independence of each hernia occurrence, and a fixed number of hernias, e.g. 1,852 in the present case, result in the so-called "Binomial process". This is the same as the process of observing heads (or tails) in a coin tossing trial with 1,852 tosses. Let us focus on the right-sided hernias - as we, for example, might focus on the "heads" outcome in coin tossing - and ignore any systematic factors for now. Using the above observed proportion 0.549 as the estimated probability of observing a right sided hernia for any given hernia occurrence, we can fully specify our statistical model for the occurrence of right-sided hernias in a sample of 1,852 hernias, in which no systematic factors are taken into account, as a "Binomial

process with probability 0.549 in 1,852 observations".[12]

But if the systematic factors age and gender are taken into account, the statistical model is slightly more complicated. This is because the inclusion of systematic variation in the model goes into the probability of observing a right-sided hernia. That is, in the model with age and gender as systematic factors, the probability of observing a right-sided hernia is no longer a constant with the estimated value 0.549, but a function of age and gender, or varies by age and gender. The precise functional form of the relationship between the probability of observing a right-sided hernia, and age and gender, will not concern us here (but is usually taken to be a logistic function). A patient therefore has a different chance of getting a right-sided hernia from that of another patient, if both are of different age and gender.

We now come to the main point of the above seemingly long-winded digression. If it were true that an appropriate model of the occurrence of the right-sided of hernias in a group of patients is a sequence of independent Bernoulli processes, with a probability systematically varying by age and gender, then the residual variation in the data (using an appropriate definition of "residual") can be considered approximately random if it is symmetric about zero with no clear systematic pattern.[13] That is, the graphical form of the residual variation should be similar to that of Figure 3. Using the hernia data set, we fitted such a Binomial model using age and gender as systematic factors, and plotted the residuals (in this example the "deviance" residuals) against a linear combination of age and gender. There is indeed a symmetric scatter about the value zero (Figure 4), showing that the residual variation can be considered random in this example.

The "demonstration" of a random residual variation does not imply that other systematic factors are unimportant in explaining the residual variation. There may well be other important factors - for example, racial and genetic factors - but we have not considered them. By focusing only on age and gender, and ignoring the rest, we have a simple statistical model which may be easy to memorize, instead of a complex multi-factor model that no one can memorize or use in every day surgical practice. The fact that the residual variation is approximately random, despite a simple two-factor model, may imply that the effects of other

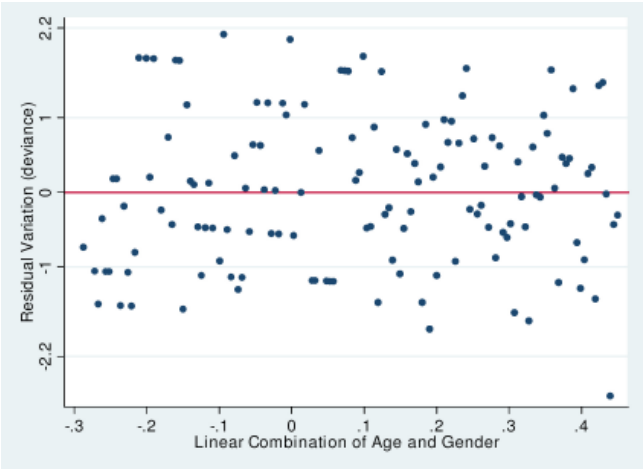

Figure 4 Residual variation plotted against systematic factors: from a real hernia data set with 1,852 observations. The residuals are symmetrically distributed about the zero value (horizontal line) with no clear systematic pattern; this is similar in form to Figure 3.

important systematic factors have approximately cancelled each other out on the average. Because we considered only two systematic factors, there will be a large residual variation. Whether random or not, such a large residual variation may render the predictions, based on only two factors, unreliable. That is, for example, even if we know that a 68 year-old male hernia patient will more likely have a right-sided hernia, we cannot predict confidently - before physical examination - that the patient will have a right-sided hernia.

Note that the precise formulation of statistical models and the fitting of models to the data is as much an art form as a "science".[14]

Randomness and unpredictability

Another conceptual mistake is to define randomness as being unpredictable. In a sense, this is true, but in another, very important sense, it is clearly misleading. Individually - i.e., in each patient - the prediction of the side of first hernia occurrence is often wrong, and in this sense randomness is unpredictable. However, as a group and given enough prior information, the proportion of patients who will have a right-sided hernia is highly predictable, becoming more reliable as the number of patients in that group increases. For example, it is difficult to predict whether an individual patient will have a right-

sided hernia, given the above data from which the probability of a right-sided hernia is estimated to be 0.549. But in a new series of 1000 hernia patients - randomly drawn from the same population - we can be reasonably confident in predicting that approximately 550 will be right-sided. In another sense, although the individual occurrence of the side of hernia cannot be reliably predicted, no matter how much information is gathered, the probability of such an occurrence can be so predicted (or "estimated") based on a suitable model. Whether these probability calculations are realistic or not is another matter, however. So there are "laws of chance" even in random variation.[15]

Conclusion

From the above discussion, under any statistical model of the medical observation, we can separate the observation into two parts: systematic and random. Researchers aim to discover and explain the systematic variation, since this variation is predictable given knowledge of the associated factors. However, the knowledge of the appropriate form of random variation is extremely important in the process as well, since it is used to determine the statistical model of the data. Randomness should not be equated with "no cause", and nor should randomness imply a complete lack of predictability.

In subsequent articles, we will discuss in detail how the ideas of sources of variation and randomness are used in statistical modeling and analysis of medical research data.

REFERENCES

1. Fisher LD, van Belle G. Biostatistics: a methodology for the health sciences. New York: John Wiley & Sons; 1993. p. 1-16.
2. Machin D, Campbell MJ. Design of studies for medical research. Chichester: John Wiley & Sons; 2005. p. 9-12.
3. The notion that a single, possibly unique, observation can be assigned a probability measure is not universally accepted. Nonetheless, for a statistical theory compatible with such a view of probability, see: Savage LJ. The foundation of statistics. New York: Dover Publications Inc.; 1972.
4. Tomassi P. Logic and scientific method. In: Phillips CI, editor. Logic in medicine. 2nd ed. London: BMJ Publishing Group; 1995. p. 39-43.
5. Thagard P. How scientists explain disease. Princeton: Princeton University Press; 1999. p. 181-98.
6. Hainault P, Wiman KG. 30 years and a long way into p53 research. Lancet Oncol 2009;10:913-9.
7. For example, see the classic frequentist definition of randomness in: von Mises R. Probability, statistics and truth. New York: Dover Publications Inc.; 1957. p. 23-9.
8. Murphy EA. The logic of medicine. 2nd ed. Baltimore: The Johns Hopkins University Press; 1997. p. 92-5.
9. Ruhla C. The physics of chance. Oxford: Oxford University Press; 1992. p. 66-125.
10. Gillispie CC. Pierre-Simon Laplace, 1749 - 1827: a life in exact science. Princeton: Princeton University Press; 1997.
11. Stigler SM. The history of statistics: the measurement of uncertainty before 1900. Cambridge, Massachusetts: Belknap Press; 1986. p. 62-98.
12. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford: Blackwell Scientific Publications; 1994. p. 55-9.
13. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. London: Chapman & Hall; 1989. p. 37-40.
14. Chatfield C. Problem solving: a statistician's guide. London: Chapman & Hall, 1988.
15. Levinson HC. Chance, luck and statistics. New York: Dover Publications Inc; 1963. p. 218-42.