Special Article

The Enigma of Statistics and Modern Cardiology

HUGO ECTOR

University Hospital Gasthuisberg, Leuven, Belgium

Key words:
Statistical error, confidence intervals, correlation, statistical power, clinical trials, logistic regression.

Manuscript received:
July 22, 2004;
Accepted:
August 10, 2004.

Address:
Hugo Ector

Cardiology Department
University Hospital
Gasthuisberg
Herestraat 49
B-3000 Leuven,
Belgium
e-mail:
Hugo.Ector@med.kuleuven.ac.be

I still remember the day, in the year 1966, when I received a present from my mentor and later Chief of Cardiology, Prof. Dr. Hilaire De Geest. It was the book "Elementary statistics with applications in medicine and the biological sciences" by Frederick E. Croxton.[1] For me, this book was the start of pursuing an understanding of statistics in daily life and in medical practice. It was the first volume in a long line of books.

In his introduction, Croxton maintains that "nearly everyone involved in any aspect of medicine needs to have some knowledge of statistics." The reality is that for many clinicians, statistics is limited to a "$p<0.05$ = ok." I do not blame my colleagues who skip the paragraph on statistical methods. They have never had the opportunity to learn from concise and clear descriptions of the key features. In the years since 1966, I have experienced how some authors can describe difficult methods in a clearly understandable language. Others fail completely.

As a teacher, I tell my students that life is impossible without a basic knowledge of statistics. My feelings on the subject have resulted in an annual seminar of 90 minutes. This article is the summary of that seminar. It is a summary and a transcription of some of the best pages I have managed to find.

Statistical errors

Type 1 and type 2 errors, due to their probabilistic nature, are discussed in every textbook of statistics. I learned about type 3 and type 4 in a paper by Robin and Lewiston.[2]

- **Type 1** (alpha) errors are false positive errors that assign statistical benefit to a given modality, when such benefit does not exist. When we receive the famous message "$p<0.05$" that treatment A is better than treatment B, we understand that the risk of a false positive result is less than 5/100! When we have a "$p<0.05$", we accept that the result is "statistically significant".
- **Type 2** (beta) errors are false negative errors: the observed distinction is regarded as not significant (no benefit), whereas in reality a benefit exists. Type 2 errors can arise in circumstances in which the numbers of observations are too small: insufficient power of the study. The classical design of a trial requires a maximum risk of a type 2 error of 0.2 (i.e. 20/100), or, in other words, the chance of detecting a real difference (power of the study) must be at least 0.8 (80/100).
- **Type 3** errors are errors in which the risk of a given medical or public health approach is underestimated, undetected or not specifically sought, leading to an underestimate of the risk-benefit balance.
- **Type 4** errors arise because the risks of a given medical intervention are overestimated, leading to under-use or abandonment of a useful intervention.

It should be emphasised that type 3 errors appear to be much more frequent than type 4 errors. This stems, in part, from the failure of medicine to establish effective mechanisms for the rapid recognition and correction of errors. Most recorded medical history deals with the triumphs of medicine. The disasters are usually interred along with their patient victims.[2]

Approximate 95% confidence limits of a proportion P expressed as a percentage (%)

$$P \pm 1.96\sqrt{\frac{P(100-P)}{N}}$$ N = number of observations

When we take a sample from the entire population, we are approximately 95% certain that the true percentage for this population will be between these confidence limits. This approximation works well for big sample sizes.

When we want 99% certainty (99% confidence limits) we have to use a factor of 2.57 instead of 1.96.

Approximate 95% confidence limits of the difference between proportions P1 and P2 expressed as percentages (%)

P_1 and P_2 are the outcomes in % in two independent groups

$$(P_1 - P_2) \pm 1.96\sqrt{\frac{P_1(100-P_1)}{N_1} + \frac{P_2(100-P_2)}{N_2}}$$

N_1 , N_2 = number of observations in group 1, 2

The difference is significant at the 5% level when the 95% confidence limits of this difference do not include 0. The risk of a type 1 error is less than 5%.

Again, when we want 99% certainty (99% confidence limits) we have to use a factor of 2.57 instead of 1.96.

"Statistics with confidence[3]" is a must in the library of every clinician.

Correlation of two variables[4]

Correlation coefficients quantify the relationship between two continuous variables.

The correlation coefficient "**r**" is calculated with the help of a statistical package. If we square the correlation coefficient, we obtain the coefficient of determination ($\mathbf{r^2}$), which can be interpreted as

$$r^2 = \frac{\text{Explained Variation}}{\text{Total Variation}}$$

A correlation of 0.4 means that we can explain 16% (0.4^2=0.16, i.e. 16%) of the total variation.

This basic idea holds true for two-variable linear or non-linear correlation and for multiple correlation. The coefficient of determination $\mathbf{r^2}$ tells us the proportion of the variation present in the dependent variable that has been explained by the use of the estimating equation. The statistical package will give a value of "**r**" and also a "**p**" value (risk of type 1 error: risk that the actual correlation is zero). A p value of <0.05 is considered as statistically significant.

The presence of a correlation between two sets of data does not necessarily mean that causation is present, even though the correlation may be high.

Size and power of a clinical trial[5]

A common statistical approach is to focus on a single outcome of patient response that is dichotomous: namely, each patient's outcome on treatment can be classified either as a "success" or "failure" (e.g. death in a year = failure, survival = success).

One has to decide on the following items:

- P_1 = percentage of successes expected from one treatment (usually the standard)
- P_2 = percentage of successes from the other treatment that one desires to detect as being different from P_1
- alpha (α) = type 1 error = the risk of a false-positive result
- beta (β) = type 2 error = the risk of a false-negative result
- $1-\beta$ = the power to detect a real difference of magnitude (P_1-P_2) = the degree of certainty that the difference, if present, would be detected.

P_1 and P_2 are theoretical percentage successes for the two treatments, which might be achieved if each were given to a large population of patients.

Example: Let us choose P_1 = 90% for placebo; P_2 = 95% for new treatment; α = 0.05 (we accept 5% risk of a false-positive result); β = 0.1 (we allow 10% risk of a false-negative result, in other words we require a power of 90%).

The required number of patients on each treatment (N) for a two-sided test is given by the following formula:

$$N = \frac{P_1(100-P_1)+P_2(100-P_2)}{(P_2-P_1)^2} \times F(\alpha,\beta)$$

F(α, β) is a function of α and β:
For α = 0.05 and β=0.2, F = 7.9 (Power = 80%)
For α = 0.05 and β=0.1, F = 10.5 (Power = 90%)

For our example we calculate:

$$N = \frac{90\times10+95\times5}{(95-90)^2} \times 10.5 =$$

= 578 patients required on each treatment (total number 1156)

Risk reduction, statistical significance, clinical relevance

It is clear that statistical significance is not synonymous with clinical relevance. However, in many papers, most emphasis is given to the relative risk reduction or to the odds ratio.

This obscures the clinical relevance, which is best reflected in the absolute risk reduction. When we reduce a risk from 50% to 25% or from 2% to 1%, in both instances, we obtain a relative risk reduction of 50%. However, the difference in clinical relevance is reflected in the absolute risk reduction: 25% versus 1%. In order to save one patient, the number-needed-to-treat is 4 (=100/25) versus 100 (=100/1).[6]

The number of patients who have to be treated to prevent one outcome event is directly related to the cost-benefit ratio of a treatment. The number-needed-to-treat is calculated by taking the reciprocal of the absolute risk reduction.

Association and prediction

Clinicians want to make predictions about outcomes for individual patients. Especially for diagnostic tests, positive predictive values are needed. I am always surprised that in most papers confidence limits for positive and negative predictive values are omitted.

It may escape our attention that the more infrequent an event, the higher will be the negative predictive value of any test. In such instances, this parameter does not reveal anything about the quality of the test. On the other hand, the positive predictive value is often not 50%, thus no better than the result of tossing a coin.

Prediction is very often confounded with association. Between some parameters and an outcome there can be a significant association, while positive predictive values remain poor. I refer to some discussions on this issue.[7-10]

Trials, observational studies, data dredging

For me it was an encounter with Prof. Dr. Alvan R. Feinstein in 1970 that introduced me to the concept of "data dredging."[11] A large number of statistical associations can be explored in an automated manner for diverse individual groups, agents, and outcomes. Whenever a "statistically significant" result emerges during the myriads of computations, the event may be proposed as a cause-effect relationship. When we examine 100 correlations, at least 5 will turn out to be significant!

We live in an era where clinical trials tend to dictate the law. What the physician believes, thinks, suspects, or has a hunch about, is assigned to "the not known" category. "Knowing" is defined on the basis of an arbitrary but accepted statistical test performed in a randomised clinical trial.[12] "We owe patients involved in the assessment of new therapies the best that science and ethics can deliver. Today, for most unproved treatments, that is a properly performed randomised clinical trial."[13]

However, because of many limitations, most of our therapeutic decisions are not entirely the result of evidence based medicine.[14] Randomised trials are infeasible for studying: (i) multiple therapeutic candidates; (ii) "instabilities" due to rapid technologic improvements; (iii) long-term adverse effects.[15] The basic science of patient care will also require our major attention to the events and observations that occur in the ordinary circumstances of clinical practice. Valuable observational studies remain necessary.[15] Increased attention is needed to improve the basic science of patient-oriented research (clinical epidemiology), and to evaluate causal relations more rigorously.[16] A remarkable conclusion was reached by Vandenbroucke.[17] Restriction of research topics, design, and analysis helps observational research to attain the desired benefits of randomisation, and gives observational research the chance to be as credible as randomised evidence.

Logistic regression

I have met few people who are aware of the basics of logistic regression. I found the best description, with an example, in an SPSS manual.[18]

Predicting whether an event will or will not occur, as well as identifying the variables useful in making the prediction, is important.

In ***multiple linear regression***, the regression coefficient tells you the amount of change in the dependent variable for a one-unit change in the independent variable.

The logistic model is rewritten in terms of the odds of an event occurring. The odds of an event occurring are defined as the ratio of the probability that it will occur {Prob(event)} to the probability that it will not {Prob(no event)}.

$$\frac{\text{Prob(event)}}{\text{Prob(no event)}} = e^{Z} \quad \log\left(\frac{\text{Prob(event)}}{\text{Prob(no event)}}\right) = Z$$

$$Z = B_0 + B_1X_1 + B_2X_2 + \ldots B_PX_P$$

Prob(no event) = 1 – Prob(event)
e is the base of the natural logarithms, approximately 2.718
B_0, B_1, B_2...are coefficients estimated from the data
X_1, X_2... are the independent variables

The logistic coefficient can be interpreted as the change in the log odds associated with a one-unit change in the independent variable.

It is clear that a statement based on this ingenious concept must offer more than a phrase such as: "logistic regression has demonstrated a statistically significant correlation between ... ." For the sake of their readers, reviewers should require specification of Prob(event) for typical values of the independent variables. Such data remain absent in most papers and hide the "absolute" truth.

The Kaplan-Meier principle in daily life

The Kaplan-Meier technique is used for the calculation of survival proportions or the so-called "free-of-event proportions". A concise and easy-to-understand explanation is given by Machin and Garner.[19]

A simple application for daily life is the estimation of a "free-of-event proportion" by hand. Suppose the risk of an event is 5% (0.05). For a constant risk of 0.05, after one year the chance of a free-of-event is 0.95, after two years 0.95^2, after three years 0.95^3 ...

Epilogue

In this article, I have only made up a bunch of other men's flowers, providing of my own only the string to tie them together.[20] I feel obliged to refer to two more statistical flowers: one for the beginner[21] and one for advanced study.[22] One flower can embellish the infirmary. Having a good memory for a few statistical principles enlightens our clinical work and duty.

References

1. Croxton FE: Elementary statistics with applications in medicine and the biological sciences: Dover Publications, Inc. New York, 1959.
2. Robin ED, Lewiston NJ: Type 3 and type 4 errors in the statistical evaluation of clinical trials. Chest 1990; 98: 463-465.
3. Altman DG, Machin D, Bryant TN, Gardner MJ: Statistics with confidence, 2nd edition. BMJ Books, 2000.
4. Croxton FE: Elementary statistics with applications in medicine and the biological sciences: Dover Publications, Inc. New York, 1959; pp 109-147.
5. Pocock JS: Clinical trials: A practical approach. John Wiley & Sons, New York, 1983; pp 123-125.
6. de Craen AJM, Vickers AJ, Tyssen JGP, Kleynen J: Number-needed-to-treat and placebo controlled trials. Lancet 1998; 351: 310.
7. Drazner MH, Rame JE, Stevenson LW, Dries DL: Prognostic importance of elevated jugular venous pressure and a third heart sound in patients with heart failure. N Engl J Med 2001; 345: 574-581.
8. Ector H: Letter to the editor. N Engl J Med 2001; 345: 1912-1913.
9. Armoundas AA, Rosenbaum DS, Ruskin JN, Garan H, Cohen RJ: Prognostic significance of electrical alternans versus signal averaged electrocardiography in predicting the outcome of electrophysiological testing and arrhythmia-free survival. Heart 1998; 80: 251-256.
10. Ector H: Letter to the editor. Heart 1999; 82: 534.
11. Feinstein AR: Scientific standards in epidemiologic studies of the menace of daily life. Science 1988; 242: 1257-1263.
12. Hellman S, Hellman DS: Problems of the randomized clinical trial. N Engl J Med 1991; 324: 1585-1589.
13. Passami E: Clinical trials: are they ethical? N Engl J Med 1991; 324: 1589.
14. Kubler W: Treatment of cardiac diseases: evidence based or experience based medicine. Heart 2000; 84: 134-136.

15. Feinstein AR: The limitations of randomized trials. Ann Intern Med 1983; 99: 544-550.
16. Concato J, Horwitz RI: Beyond randomized versus observational studies. Lancet 2004; 363: 1660-1661.
17. Vandenbroucke JP: When are observational studies as credible as randomized trials? Lancet 2004; 363: 1728-1731.
18. Norusis MJ: SPSS Professional Statistics 7:5. SPSS Inc. Chicago, IL 60611, 1997; pp 37-64.
19. Altman DG, Machin D, Bryant TN, Gardner MJ: Statistics with confidence, 2nd edition. BMJ Books, 2000; pp 93-104.
20. Quotation from Michel de Montaigne, Essays, 1580:
21. Glantz SA: Primer of Statistics: McGraw-Hill Book Company, New York, 1987.
22. Feinstein A, lvan R: Multivariable Analysis: An Introduction: Yale University Press, New Haven and London, 1996.