Without statistical methods, clinical medicine would not be where it is today

An indispensable tool

Understanding of biology is the obvious cornerstone of clinical medicine. However, even when we claim to have a good mechanistic understanding of a disease, we can sometimes find that a treatment principle that should work in theory proves to have little or no effect in practice. An old and familiar example is the CAST study (1), which surprisingly showed that arrhythmia suppression with encainide and flecainide did not reduce mortality after a myocardial infarction when compared to placebo, but quite the reverse. In the field of drug development there are many Phase 3 studies that fail to show any effect of treatment, in spite of results from initial studies conducted on limited samples of healthy volunteers and patients (Phase 1 and Phase 2) showing a positive effect of the active ingredient (2, 3). In other words, a mechanistic understanding is often insufficient, most likely because we understand only a tiny part of the whole picture. In this situation, statistical methods have proven to be indispensable. When conducting randomised intervention studies, the methods of analysis do not even need to be particularly complicated.

In this issue of the Journal of the Norwegian Medical Association there are three articles that each in its own way addresses the topic of medical statistics. They illustrate that statistics have become an indispensable tool in medical research (4, 5), and that probability and statistics are part of everyday clinical practice (6). Our increasing insight into biological mechanisms has not necessarily made it any easier to understand diseases, identify risk factors or develop new treatment strategies. When complexity increases, it becomes evident that statistical methods are necessary to understand and explain the issue at hand.

As we know, medical research is largely based on drawing conclusions with the aid of p-values. Unless the p-values are misused and overinterpreted, this is a fully acceptable approach that (in combination with estimates of the effect size with an appurtenant confidence interval) has proven to be of major practical benefit, especially in analyses of experimental studies. However, as Pripp (4) points out in his article, for many the p-value is an oracular answer, the interpretation of which requires special competence. Even simple analyses may lead to erroneous conclusions, such as when multiple significance tests are undertaken.

In experimental studies that are based on randomisation and a sensible trial design, it will often be sufficient to use elementary statistical methods to undertake comparisons and estimate effects. Assigning patients randomly to different treatment groups implicitly ensures that any observed differences are caused either by an actual difference in effect or random variation. The p-values can thereby be interpreted directly.

Observational studies often pose greater challenges, since the results can be distorted by systematic bias or confounding (background) factors that have not been measured. Identifying causal associations can thus prove difficult. The gold standard for establishing causality is the randomised study, but in recent years *causal inference* has emerged as an important research area, and so-called directed acyclic graphs (DAG) are increasingly used in epidemiological research. Such graphs are useful for demonstrating associations and as tools for selecting the variables that should be included in a regression model, but they may often become highly complex and will obviously not be any better than our (occasionally limited) understanding of associations and mechanisms. The article by Stensrud and Aalen (5) clearly shows that causality is a challenging issue.

The choice between frequentist and Bayesian methods is subject to recurring debate. The frequentist tradition, which is clearly prevalent in medical science, is based on the formulation of a null hypothesis (the claim that we wish to disprove) and uses the data to reject it if the p-value is low.

If we are unable to undertake randomised studies and it is difficult to obtain new data on a sufficient number of patients, a Bayesian approach can be an alternative. Bayesian statistics summarise what we know (or assume) in advance regarding an unknown parameter in a so-called a priori distribution. The distribution of the new data from the trial is collated with this prior knowledge in an a posteriori distribution, and the likelihood of a hypothesis can then be quantified, because we estimate «the likelihood of the hypothesis given the data», and not «the likelihood of the data given the hypothesis» as in a frequentist approach. This is intuitively attractive, and as shown by Brakedal (6), Bayesian ideas are used more or less unconsciously for diagnostic purposes.

One reason why Bayesian methods have been less used than frequentist ones in medical research is that they are computationally demanding and were nearly inapplicable in practice before the advent of powerful computers. Another and even more important reason for their limited use in clinical research is that in experimental situations we do *not* want to be affected by pre-defined assumptions or beliefs. For example, pharmaceutical companies tend to be overly optimistic with regard to the efficacy of a new drug, but unfortunately, randomised studies will often show that in reality, the efficacy is not as convincing as expected – or even absent. To prevent overly positive assumptions or beliefs from influencing the conclusions it seems most appropriate to use a frequentist approach. When using a Bayesian approach it is, however, possible to enter a so-called non-informative a priori distribution. Then, the estimation will be made on the basis of the data alone, and the conclusion will thus be the same as when traditional frequentist methods are used.

Irrespective of approach, the increasing importance of statistics in all sectors of medicine is striking. Modelling of biological mechanisms that explain disease, as well as the effect of prevention and treatment, is increasing in complexity, and conclusions can rarely be drawn from a mere mechanistic understanding of biology. Thus, medical research is often completely reliant on advanced statistical analysis.

Eva Skovlund
eva.skovlund@ntnu.no

Eva Skovlund (born 1959), professor at the Department of Public Health and General Practice, Norwegian University of Science and Technology. The author has completed the ICMJE form and declares no conflicts of interest.

References
1. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. N Engl J Med 1989; 321: 406–12.
2. Arrowsmith J. Trial watch: phase III and submission failures: 2007–2010. Nat Rev Drug Discov 2011; 10: 87.
3. Hay M, Thomas DW, Craighead JL et al. Clinical development success rates for investigational drugs. Nat Biotechnol 2014; 32: 40–51.
4. Pripp AH. Hvorfor p-verdien er signifikant. Tidsskr Nor Legeforen 2015; 135: 1462–4.
5. Stensrud MJ, Aalen OO. Hva kan vi si om kausalitet? Tidsskr Nor Legeforen 2015; 135: 1465–7.
6. Brakedal B. En kliniker og en bayesianer. Tidsskr Nor Legeforen 2015; 135: 1468–70.