Cancer Chemother Pharmacol (2001) 47 (Suppl): S35–S37

GUEST EDITORIAL

Berthold Schneider

Analysis of therapeutic efficacy in observational cohort studies

Introduction

A drug is effective in the treatment of a disease if more patients are healed, or obtain relief, with the drug than without. The proof of therapeutic efficacy requires a comparison in representative patient populations of the success of treatment with the drug with that achieved without the drug or other treatments. The standard method for the comparison of therapeutic treatments is the randomized controlled trial (RCT) as proposed in the classic books of Martini [8] and Hill [7]. In the RCT a representative sample of patients suffering from the disease of interest is included in the trial according to clear inclusion and exclusion criteria. The sample is split into two groups: one group of the included patients (usually half of them) is treated with the drug whose efficacy is to be determined (test group) while the other is treated with a control treatment (control group) which should be a placebo if no standard treatment is known for the disease. The patients are randomly assigned to each of the treatment groups. The randomization process assures equivalence of both treatment groups with respect to all individual baseline conditions, e.g. age, history, disease state etc. In statistical terms the random distributions of these conditions are identical for both groups which allows an unbiased comparison of the treatments. Guidelines for the design, performance, analysis and publication of results have been published e.g. by the European Agency for the Evaluation of Medicinal Products (EMEA).

This work was presented at Anwendungen der Biometrie in Medizin, Landwirtschaft und Mikrobiologie. Kolloquium anlässlich des zehnjährigen Bestehens der BioMath GmbH, Rostock, Germany, 16 November 2000

B. Schneider
Institute of Biometrics, Medical University Hannover,
Konstanty-Gutschow-Strasse 8,
30625 Hannover, Germany
Tel.: +49-511-5326507; Fax: +49-511-5324295

For new drugs the RCT is the only feasible method of proving efficacy. In contrast, where a drug has been on the market for a number of years without the benefit of an RCT to prove efficacy, it is desirable to use the body of knowledge that has accumulated during the practical application of the drug to determine its efficacy and safety. Observational epidemiological studies are valid methods of providing evidence of efficacy and safety of drugs and are sanctioned by the European Commission [4]. A basic requirement is that the study data allow an unbiased comparison of the results observed after application of the test drug with the results observed after other treatments (or after no treatment). In epidemiological cohort studies a representative set (cohort) of individuals is sampled from the population of interest (patients suffering from a certain disease) and the applied treatments and observed responses are documented together with all other relevant patient and treatment characteristics (so-called covariates). The treatment is left to the decision of the physician or patient and is not affected by the study. Clearly, in the population (and also in the selected cohort) patients will be taking the drug under investigation concomitantly with other drugs or treatments. Data collection can be done prolectively for new patients coming to the practice or hospital or retrolectively from the files of formerly treated patients. The term 'retrolective' was introduced by Feinstein [5] to distinguish between the mode of data acquisition (which may be retrolective or prolective) and the mode of treatment (which may be retrospective or prospective). Whatever method is employed the results of the study depend entirely upon the quality of the data and their collection. Guidelines for 'good epidemiological practice' have recently been published [6].

Improved concept of retrolective cohort studies

Retrolective cohort studies are particularly suited to proving the efficacy of drugs that have been in use for a long period of time. Large numbers of patients may have

been treated in the past with these drugs as well as with other medications and the medical records provide a valuable source of data. An improved methodology for retrolective cohort studies has been developed recently (Retrospect, IFAG, Basle) [12] featuring a number of aspects.

Prior to study initiation a detailed study protocol should be developed and approved by all responsible persons (study leader, study organizer, biometrician, sponsor). The protocol contains details of the objectives of the study, its design, selection of centres, sample size, inclusion and exclusion criteria, data extraction, data quality control, biometric analysis and publication policy. Due to the retrolective character of the study, approval by an ethics committee is not necessary.

From centres (practices, hospitals) which use the test drug as well as other treatments a representative sample is selected and the medical practitioners invited to participate in the study. Within the centres the data relating to all patients who were treated for the disease of interest over a definite time period are extracted from the medical records. The selection of patient documents and data are determined by the inclusion and exclusion criteria laid down in the study protocol, and by standard operation procedures (SOPs). Information about the centres from structured doctor interviews from other institutions is included, if necessary.

The extracted data are coded and transferred to standardized case report forms (CRF) after anonymization. Rules for coding and transfer of data to the CRF should be detailed in the study protocol. Where possible, international standard codings, e.g. for diagnoses (ICD), drugs, medical procedures (ICMP) or unwanted events (WHO coding) should be used. An alternative to the manual transfer of data to the CRF is online remote data entry into computer networks. In the absence of remote data entry, the data in the CRF are entered into a database on a computer system and checked for completeness, correctness and plausibility. If necessary double data entry may be performed. Data quality control and correction should be performed by independent monitors. Finally, a quality assurance audit performed by individuals not associated with the study can be undertaken comparing the data in a random sample of the original patient files with those held on the database.

The biometric analysis and publication of the results should be performed according to the procedures stated in the study protocol. In particular, the primary and secondary outcome variables (responses) and the test and control groups must be specified in the protocol. Changes in procedures or additional analyses are laid down in amendments to the protocol.

Biometric analysis

In efficacy analysis of drugs the primary aim is to compare the response (defined as primary outcome variable) of patients treated with the test drug with the response of those receiving the control treatment. In the RCT comparability of test and control groups is achieved by randomization of treatment allocation. In observational studies treatments are allocated by the physician or patient, and these decisions may depend on factors which may also influence the treatment response. Therefore an immediate comparison of test and control groups may be biased by these factors. To obtain unbiased results, an adjustment for these factors is necessary before the treatments are compared. Biometric analysis provides two equivalent approaches for such an adjustment: stratification (subgroup analysis) and regression techniques (covariate analysis).

In the stratification approach analyses are performed on the subgroups (also called strata) which are homogeneous in relation to the influencing factors (also called covariates). Unbiased comparisons of treatments can be performed within the subgroups (strata) and the results of each of the comparisons are pooled to get an overall result. A special case of stratification is the matched pairs technique, where each patient treated with the test drug is matched with a patient receiving the control treatment whose factors are as similar as possible to those of the test patient. Each of these pairs form a subgroup or stratum.

With the regression technique the influence of the factors on the response is modelled by an appropriate regression function which is estimated from the data of the study irrespective of the applied treatment. With this function each individual response is adjusted to unique factor reference values (e.g. to the factor means in the study population) and the adjusted responses are compared between treatment groups.

The effectiveness of both techniques depends on the completeness of the included factors, and in the case of the regression approach, on a suitable choice of regression function. Only if all factors that have a relevant influence on the response are included in the analysis can a satisfactory adjustment be expected. The influencing factors include not only patient characteristics such as age, gender, anamnestic data, pretreatment, response to pretreatment, state of the disease at start of treatment, additional diseases, concomitant treatments etc, but also characteristics of the treatment centre such as the age or specialization of the treating doctor. Clearly, a large number of factors have to be considered when performing stratification or regression analyses and this leads to practical difficulties. For example, in a case in which ten binary (i.e. present/not present) factors have to be taken into account there would be 1024 combinations of these factors for which stratification must be performed, and this is essentially impractical.

A solution to this dilemma is the introduction of a balancing score. This is a function of all factors that influence the response. If such a score can be determined, it is only necessary to adjust for the values of this function either by stratification or the regression technique. A special balancing score is the propensity score

introduced by Rosenbaum and Rubin [10]. This score is the probability of allocating a patient to the test drug as a function of his or her individual factors (characteristics) and the characteristics of the treatment centre. This probability can be estimated with the data from the study, e.g. by logistic regression. These authors showed that with the propensity score an optimal adjustment is possible and an unbiased comparison of treatments can be achieved if the propensity score takes into account all the relevant factors that influence the response and the treatment allocation. In addition, the propensity score provides a valuable insight into the conditions and structures of treatment decisions in practice. This in itself is a very important result of observational cohort study analysis.

Conclusions

The practicability and effectiveness of adjustment with the propensity score has been demonstrated in many observational studies [3, 9, 11]. These studies have shown that the propensity score can provide a very good adjustment of the inhomogeneities caused by numerous influencing factors and that a valid and unbiased comparison of treatments is possible with observational studies. Extensive comparisons of RCT and observational studies for a variety of diseases have shown that the estimated effects of treatment are consistent for both types of studies. In conclusion, the following quotation is particularly relevant: "Our results suggest that observational studies usually do provide valid information. They could be used to exploit the many recently developed, clinically rich databases. Only with a greater willingness to analyse these databases is it possible to achieve a realistic understanding of how observational studies can best be used" [1, 2].

References

1. Benson K, Hartz AJ (2000) A comparison of observational studies and randomized, controlled trials [see comments]. New Engl J Med 342: 1878
2. Concato J, Shah N, Horwitz RI (2000) Randomized, controlled trials, observational studies, and the hierarchy of research designs [see comments]. New Engl J Med 342: 1887
3. D'Agostino RBJ (1998) Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 17: 2265
4. European Commission (1999) Directive 1999/83/EC of 8 September 1999. Official Journal L 243/9: 0009–0011
5. Feinstein AR (1985) Clinical epidemiology. Saunders, Philadelphia
6. German Professional Society of Epidemiology (2000) Guidelines and recommendations for insuring good epidemiological practice (GEP). Gesundheitswesen 62: 295
7. Hill AB (1937) Principles of medical statistics. Lancet, London
8. Martini P (1932) Methodenlehre der therapeutischen Untersuchung. Springer Verlag, Berlin
9. Perkins SM, Tu W, Underhill MG, Zhou X-H, Murray MD (2000) The use of propensity scores in pharmacoepidemiologic research. Pharmacoepidemiol Drug Safety 9: 93
10. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70: 41
11. Rubin DB (1997) Estimating causal effects from large data sets using propensity scores. Ann Intern Med 127: 757
12. Wittenborg A, Bock PR, Hanisch J, Saller R, Schneider B (2000) Vergleichende epidemiologische Studie bei Erkrankungen des rheumatischen Formenkreises am Beispiel der Therapie mit nichtsteroidalen Antiphlogistika versus einem oralen Enzymkombinationspräparat (Comparative epidemiological study in patients with rheumatic diseases illustrated in an example of a treatment with non-steroidal anti-inflammatory drugs versus an oral enzyme combination). Arzneimittelforschung 50 (II): 728