

**ON THE USE OF STATISTICAL EXPERT SYSTEMS
FOR SUPPORTING QUALITY ASSURANCE IN ANALYSIS OF CLINICAL TRIALS**

Knut M. Wittkowski

Dept. Medical Biometry, Eberhard-Karls-University
Westbahnhofstr. 55, D-7400 Tübingen, Fed. Rep. Germany

ABSTRACT

Statistical analysis of clinical trials is predetermined by the study protocol. Because this knowledge cannot be represented in commercially available analysis systems (e.g. BMDP, P-STAT, SPSS, SAS), their application is susceptible to mistakes which are often overlooked. A knowledge based front end for the analysis of clinical trials is presented which consists of six layers corresponding to different levels of abstraction in the knowledge. It is demonstrated that this "expert system" can improve statistical analysis in that access to data and programs is facilitated, the probability of erroneous applications of analysis programs is reduced, and the user is assisted in interpreting both confirmatory and exploratory results. This approach leads also to improved standards for the analysis of clinical trials.

Key words: software package, experimental design, confirmatory analysis

ZUSAMMENFASSUNG

Die Auswertung klinischer Studien ist im Versuchsprotokoll bereits weitgehend festgelegt. Da dieses Wissen in den kommerziell verfügbaren Systemen (z.B. BMDP, P-STAT, SPSS, SAS) nicht gespeichert werden kann, ergeben sich bei der Anwendung leicht Fehler, die zudem oft nicht bemerkt werden. Es wird eine wissensbasierte Schnittstelle zur Auswertung klinischer Studien vorgestellt, die aus sechs Schichten - entsprechend unterschiedlichen Abstraktionsebenen - aufgebaut ist. Dieses "Expertensystem" kann die Qualität statistischer Auswertungen verbessern, indem es den Zugriff auf Daten und Programme erleichtert, die Wahrscheinlichkeit fehlerhafter Anwendung von programmen reduziert und den Benutzer bei der Interpretation confirmatorischer und explorativer Resultate unterstützt. Dieser Ansatz führt zudem auch zu einer sinnvollen Standardisierung bei der Auswertung klinischer Studien.

Schlüsselworte: Auswertungssystem, Versuchsplan, confirmative Statistik

Introduction

Clinical trials differ from retrospective studies in that primary goal, experimental design, intended analysis methods, sample size, and randomization procedure need to be defined prior to the collection of data. In the ideal case, the study protocol contains most of the knowledge necessary for performing confirmatory analyses concerning the primary hypothesis and exploratory analyses which might lead to new hypotheses.

Commercially available analysis systems cannot store knowledge on formal relations (Wittkowski 1984). Thus the user has to extract all relevant information from the study protocol and to translate it into the appropriate control language or he must draw the necessary conclusions himself with each subsequent call of a procedure, e.g. selection of data subsets, definition of derived variables, graphical and tabular representation of data, and application of analytical methods.

Such a repeated activity is both boring and susceptible to mistakes. Either some important aspects of the model or some assumptions implicit to the methods might be overlooked so that the interpretation of treatment effects implicit to the method differs from the medical interpretation or the data are not arranged in the way expected by the program (Wittkowski 1985b, Bernhard et al. 1988). Because "p-values" do not show whether or not they are semantically meaningful, errors are often overlooked (Altman 1982, Wittkowski 1986b). Quality assurance in the analysis of clinical trials could be assisted by analysis systems that automatically consider the study protocol during each step in the design and analysis of experiments.

Where knowledge of different domains (computer science, statistics, biometry, medicine) has to be integrated, a monolithic system would be extremely difficult to construct. In the present paper, we will consider a system, which consists of several layers corresponding to the different levels of abstraction in the knowledge (Elliman, Wittkowski 1987).

Methods

The system presented is logically divided into six layers according to different levels of abstraction in the knowledge: EXECUTION, ACCESS, SEMANTICS, STRATEGY, DOMAIN, DIALOGUE (Wittkowski 1988c). Each layer contains knowledge of a different area of expertise implemented using dedicated knowledge representation techniques. The bottom EXECUTION layer contains the observed data and the available analysis programs. It may consist of commercially available statistical analysis systems.

In medical research, we are typically faced with a multitude of derived attributes and analysis programs. In the ACCESS layer we have knowledge on the tree of links between original and derived attributes (c.f. e.g. Blum 1982) and on parameters and data structures required by the programs. All requests for accessing data and programs are handled at this level. The SEMANTICS layer contains knowledge of the lattice structure (Wittkowski 1985a) of observational units (e.g. patient, time, treatment), i.e. whether or not these units are nested or crossed both in the study design and (implicitly) in the statistical methods. Physiological models and statistical concepts are represented in the STRATEGY layer. At this level, we have on the one hand theoretical relations, i.e. the distinction of factors, strata and observed variables, SI-units, level and type of scales, integrity constraints (e.g. ranges and number of possible different realizations), and the sampling strategy. On the other hand we have knowledge on statistical concepts. At the DOMAIN level, medical knowledge is either entered or generated. In the top DIALOGUE layer, a personalized knowledge of the user is constructed and all activities concerning a given experiment are monitored (THISTED 1987).

Results

Within each layer different activities of human experts (e.g. in computer science, statistics, biometry, medicine, etc.) are to some degree taken over by the expert system. In the implementation of PANOS-ES we have started with the implementation of the ACCESS, SEMANTICS, and STRATEGY layer.

Most statistical expert systems (e.g. Oldford, Peters 1986, Nelder, Wolstenholm 1986, Lubinsky, Pregibon 1987) have so far been designed for the purpose of assisting the user in fitting models to data (observed relations). It has been demonstrated (Wittkowski 1985b) that this approach is not appropriate for the analysis of clinical trials, where models must be chosen primarily with respect to knowledge which is independent of the data. Thus inference in the higher layers is based primarily on theoretical, hypothetical, and observable relations. As a consequence, all properties of the data (e.g. homoscedasticity, Gaussian distribution of residuals, symmetry of covariance matrices) can be handled by the analysis systems at the EXECUTION level. Implications of this approach on the design of distributed systems have been discussed elsewhere (Wittkowski 1984, 1985b).

Commands and parameters for underlying database management and analysis systems are generated in the ACCESS layer. A knowledge based spread-sheet editor has been developed as part of the PANOS-ES system.

It guarantees that only observed data may be edited and that corrections are propagated to all related observational units and to all derived attributes. Because both program calls and requests for updating, aggregating, and retrieving data are handled at this level (Wittkowski 1988b), data structures are automatically generated according to the programs' requirements. Common problems concerning misspecification of data structures and program parameters (c.f. Bernhard et al. 1988) in analysis of medical data are avoided.

At the SEMANTICS level appropriate data structures and methods are selected. The lattice structure of observational units is utilized e.g. to avoid generating non-observational relations and to distinguish between dependent and replicated data (Wittkowski 1988b). When the user requests computation of derived attributes, the knowledge on SI-units, scales, and ranges is considered in addition to the knowledge on observational relations to check the semantic integrity of this request and to propagate this knowledge to the derived attribute. For instance, an attribute whose range includes the origin must not be used as a denominator, even if no zeros have been observed. Theoretical relations are also considered when a method is selected. For instance, the lattice structure is used to identify crossed and nested effects and the scale types are used to choose among conditional and unconditional rank tests. Thus it is guaranteed that the methods chosen reflect the problem under consideration and that these methods are then applied to sufficient data subsets.

In the STRATEGY layer, the model is checked for consistency and primary or secondary hypotheses are formalized. Only those hypotheses will be accepted that conform to the theoretical relations. A hypothesis concerning expectation, for instance, is allowed only if the attribute is at least interval scaled, but rejected if the attribute is an ordinal ranking (e.g. level of pain or quality of life). Once the hypothesis has been accepted, the level of significance, a medically relevant alternative, and the desired power to detect this alternative hypothesis may be entered so that the sample sizes can be automatically computed even in complex designs. Because both the primary hypothesis and the structure of the complete study are known, the system can distinguish between confirmatory and exploratory analyses. In the former case, it can automatically choose multiple comparison procedures or adjust p-values to reduce the probability of misinterpreting "significant" results (Wittkowski 1986a).

The knowledge implemented in the EXECUTION, ACCESS, SEMANTICS, and STRATEGY layer has been demonstrated to be useful for medical applications, although it is independent on the domain of application. Medical knowledge available at the DOMAIN level will be useful both in building models (selecting appropriate subsets of clinically relevant

variables) and in designing experiments (providing information of "relevant" differences and risk of wrong decisions).

The level of explanation both during input of knowledge and output of results will be adjusted to the users' needs using the knowledge at the DIALOGUE level. While structured dialogue procedures have been demonstrated to be efficient for problem formulation and acquisition of knowledge, natural language may be useful at this level for teaching statistical concepts and for interpreting results to inexperienced users (Wittkowski1988a).

Discussion

A knowledge based front end to statistical analysis systems can improve analysis of clinical trials in many ways. Because the system consists of several layers, the user may choose the level of assistance. A user with less expertise in statistics may use a high-level interface that allows for fast and efficient dialogue procedures, while more experienced users may opt for more control at the cost of more complex dialogue procedures.

At the higher levels, access to data and programs is simplified. The amount of information to be entered during analysis of sub-problems is significantly reduced, because most relevant information from the study protocol is automatically considered by the system. The user formulates questions and lets the system choose a method with the appropriate implicit assumptions. This concept leads not only to improved dialogue procedures but also to fewer erroneous applications of analysis programs, because relevant assumptions are not overlooked and parameters not misspecified.

All analyses not conforming to the primary goal of the trial (e.g. because a subpopulation is selected or a different hypothesis is tested) are automatically marked as being exploratory. Where sub-hypotheses are tested (e.g. simple main effects or paired comparisons), the p-values are adjusted for all confirmatory analyses, so that the frequency of misinterpreting "significant" results is reduced. Model knowledge can also be used to explain results and thus help the user to interpret them in terms of the originally formulated question.

Finally, research in knowledge representation is also important for the acceptance of statistical methods in the field of medical research: Where statistical "expertise" is explicitly defined, it can be discussed and common misleading or wrong "heuristics" can be corrected. As a consequence the choice of a statistical strategy and thus the result of a trial will become less dependent on the subjective opinion of a single expert.

References

- ALTMAN DG (1982) Statistics in medical journals.
Statistics in Medicine **1**, 59-71.
- BERNHARD G, ALLE M, HERBOLD M, MEYERS W (1988) Investigation on the reliability of some elementary nonparametric methods in statistical analysis systems. *Statistical Software Newsletter* **19**, 19-26
- BLUM RL (1982)
Discovery and representation of causal relationships from a large time-oriented clinical database: The RX project. Berlin FRG: Springer
- BOARDMAN TJ (ed. 1986)
Computer science and statistics. Washington, DC; ASA
- DEANTONI R, LAURO N, RIZZI A (eds. 1986)
COMPSTAT SG. Heidelberg, FRG: Physica
- ELLIMAN AD, WITTKOWSKI KM (1987)
The impact of expert systems on statistical database roanagement.
Statistical Software Newsletter **13**, 14-18, 26-27
- FAULBAUM F, UEHLINGER HM (eds. 1988)
Fortschritte der Statistik-Software **1**. Stuttgart, FRG: Fischer
- HAUX R (ed. 1986)
Expert systems in statistics. Stuttgart, FRG: Fischer
- LUBINSKY D, PREGIBON D (1987) Data analysis as search.
In: PHELBS B (ed. 1987) 18-35
- NELDER JA, WOLSTENHOLME DE (1986) A front end for GLIM.
In: HAUX R (ed. 1986) 155-177
- OLDFORD RW, PETERS SC (1986)
Object-oriented data representations for statistical data analysis.
In: DEANTONI R, LAURO N, RIZZIA (eds. 1986) 301-306
- PHELBS B (ed. 1987) *Interactions in artificial intelligence and statistical methods*. Aldershot, GB: Gower
- RAFANELLI M, KLENSIN J, SVENSSON P (eds. 1988)
Proceedings of the fourth international workshop on statistical and scientific database management. (in press)
- ROGER FH, GRONROOS P, TERVO-PELIKKA R, O'MOORE R (1985 eds.)
Medical informatics europe 85. Berlin, FRG: Springer
- THISTED RA (1986) Tools for data analysis management.
In: BOARDMAN TJ (ed. 1986) 152-159
- VAN EIMEREN W, ENGELBRECHT R, FLAGLE CD (1984 eds.) *Third international conference on system science in health care*. Berlin, FRG: Springer
- WITTKOWSKI KM (1984) On the use of structural information for a statistical expert system in medical research, In: VAN EIMEREN W, ENGELBRECHT R, FLAGLE CD (1984 eds.) 1140-1143
- WITTKOWSKI KM (1985a) *Ein Expertensystem zur Datenhaltung und Methodenauswahl für statistische Anwendungen*. Stuttgart: Dissertation; Internat. Bookseller Hans Hartinger Nachf., Xantener Str. 14, D-1000 Berlin 15
- WITTKOWSKI KM (1985b)
An expert system for statistical applications in medical research.
In: ROGER FH, GRONROOS P, TERVO-PELIKKA R, O'MOORE R (1985 eds) 153-162
- WITTKOWSKI KM (1986a) An expert system for testing statistical hypotheses.
In: BOARDMAN TJ (ed. 1986), 438-443
- WITTKOWSKI KM (1986b) Generating and testing statistical hypotheses: strategies for knowledge engineering. In: HAUX R (ed. 1986) 139-154
- WITTKOWSKI KM (1987) An expert system approach for generating and testing statistical hypotheses. In: PHELPS B (ed. 1987) 45-59
- WITTKOWSKI KM (1988a) Intelligente Benutzerschnittstellen für statistische Auswertungen. In: FAULBAUM F, UEHLINGER HM (eds. 1988) 212-225
- WITTKOWSKI KM (1988b)
Knowledge based support for the management of statistical databases.
In: RAFANELLI M, KLENSIN J, SVENSSON P (eds. 1988) (in press)
- WITTKOWSKI KM (1988c) Building a statistical expert system with knowledge bases of different levels of abstraction.
In: *COMPSTAT 1988*. Heidelberg, FRG: Physica, (in press)