18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam 1996
7.4.1: Clinical Decision Making

# Representation of domain knowledge needed to define relevant study variables for clinical trials in urology

Wendy d'Hollosy, Pieter F. de Vries Robbé*, Nicolaas J.I. Mars**,
Wim P.J. Witjes, Frans M.J. Debruyne, Hessel Wijkstra

University Hospital Nijmegen, Dept. of Urology, Nijmegen, The Netherlands.
* University of Nijmegen, Dept. of Medical Informatics, Nijmegen, The Netherlands.
** University of Twente, Dept. of Computer Science, Enschede, The Netherlands.

*Abstract* - **An important aspect of clinical trial design is to define which study variables are relevant. This describes the data that have to be collected during the conduct of a trial. An underestimation of this aspect of trial design may lead to problems in the later data management process.**

**We examined the knowledge needed to determine relevant study variables. This paper describes the development of a conceptualization model, also called an *ontology* [1], to represent this knowledge. This is the first step in building a knowledge-based system as a part of PROSYS [2].**

## I. Introduction

In university hospitals research is performed to improve patient treatment. *Clinical trials* have become the most common study designs to measure the utility of new or adjusted medical treatments [3]. A clinical trial is an experimental study to evaluate the effectiveness of a treatment. This evaluation is mostly performed by comparing the observed effects of the test treatment on a group of subjects with those of a control treatment on a similar group of subjects [4].

The study question of most clinical trials is: "What is the effect of the investigated treatment(s) in a group of subjects that are suffering from the disease of interest?". This can be positive effects (*treatment goals*) as well as negative effects (*side-effects*). Data are collected, during the conduct of a trial, to answer this question.

Study variables describe the data that have to be collected. This means that the determination of relevant study variables for a new trial strongly influences the later data management process. A wrong determination of study variables may lead to problems as[5,6]: 1.Data are collected which will not be used afterwards, and 2.Data will be needed afterwards that are not collected during the execution of the trial.

Marking study variables as relevant is based on literature, experience and advice from other investigators. It requires knowledge on the medical domain investigated, existing treatment methods, and, if available, experience gained on study variables in previous comparable trials. This is a lot of knowledge that is even more growing, because of advances in biomedical research, resulting in new or supplementary diagnostic and treatment methods and new knowledge on the medical domain investigated.

*Knowledge-based systems* are useful tools to structure and manipulate knowledge to support 'intelligent' tasks. In literature, however, no such systems are described that support the definition of relevant study variables for new clinical trials. We therefore started the design of a knowledge-based system that is able to perform this task. This system will be part of PROSYS (= PROtocol design SYStem) [2].

The first step in building a knowledge-based system concerns the representation of the domain knowledge. This paper describes the development of a conceptualisation model to represent domain knowledge needed to define relevant study variables for clinical trials in urology.

## II. Materials and methods

The development of knowledge based-systems starts with a *conceptualisation* of the application area. This means that the application area is structured into a system of domain *concepts* and *relations* between these concepts [1]. Conceptualization is independent from the way the knowledge should be encoded in the computer. This means that conceptualization takes place at the *knowledge level* and not at the *symbol level* [7].

An *ontology* is an explicit specification of a conceptualisation [1]. An ontology specifies at a higher level what *classes of concepts* are introduced for the application domain and what *classes of relations* exist between these concept classes. The choice of introduced concept classes and relation classes in the ontology is crucial to the way of representation of the final domain knowledge.

We developed an ontology to model the representation of domain knowledge for our application. Literature, clinical trial protocol documents and interviews with experts were

0-7803-3811-1/97/$10.00 ©IEEE

used to investigate the application domain. Some choices had to be made during this ontology development. These choices were based on, sometimes contradictory, criteria. The most important criteria were that it should be possible to represent all needed knowledge (*expressiveness*) and that it should be possible to represent all needed knowledge with as few concept classes and relation classes as possible without losing expressiveness (*parsimony*).

## III. Results

The developed ontology consists of 9 concept classes and 10 relation classes. This ontology is displayed in figure 1. An example of domain knowledge represented according to this ontology is shown in figures 2.

## IV. Discussion

Ontologies are used to facilitate the reuse of domain knowledge in new applications and to build knowledge acquisition tools [8]. We also developed an ontology to represent domain knowledge for our application. Here the ontology was developed to obtain a conceptualisation model that can be used to built different knowledge bases for usage by the same inference mechanism. That is because the construction of the inference mechanism of our knowledge-based system will be based on the domain knowledge representation structure defined by this ontology.

For example, our application is now intended for use in the urological domain. The ontology, however, can also be used as definition to conceptualize domain knowledge of other medical fields. This means that the usage of this knowledge-based system can be extended to other medical fields by building new knowledge-bases containing knowledge on these medical fields that is structured according to this ontology. The same inference mechanism, in cooperation with these new knowledge-bases, can then be used to support the definition of relevant study variables for new clinical trials in other medical domains.

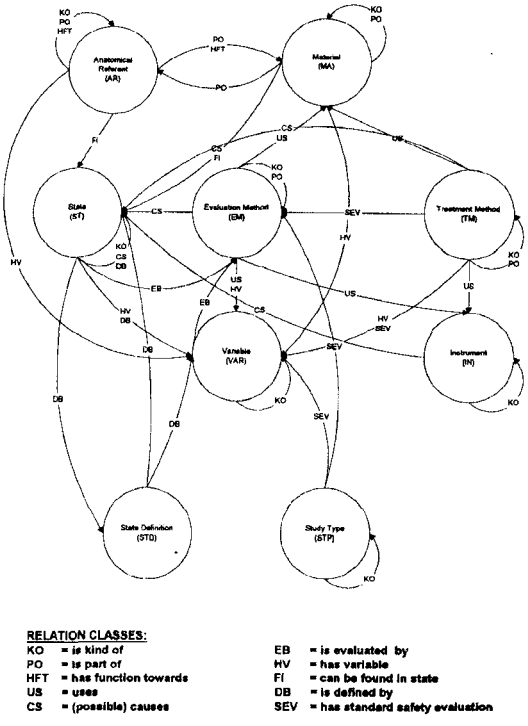


**Fig.1** Schematical representation of the developed ontology.

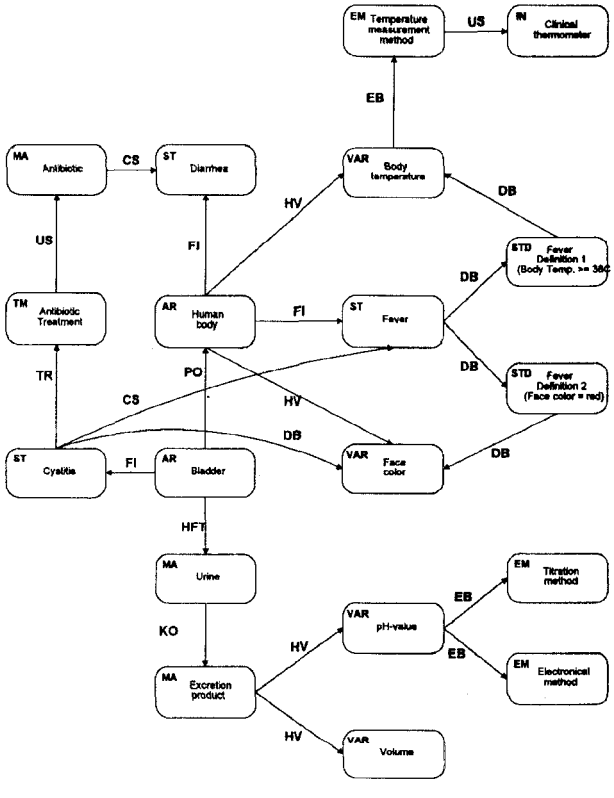


**Fig.2** Example of domain knowledge represented according to the developed ontology.

## References

[1] T.R.Gruber, "A translation approach to portable ontology specifications.", Knowledge Acquisition 5(2), pp.199-220, 1993.
[2] W.d'Hollosy, W.P.J. Witjes, N.J.I. Mars, P.F. de Vries Robbé, F.M.J.Debruyne, H.Wijkstra, " Knowledge-based support in study design of scientific urological research", Proc. IEEE-EMBS, vol. 17, 1992.
[3] The Standard of Reporting Trials Group: "A proposal for structured reporting of randomised controlled trials.", Special Communications IN: J.A.M.A. 272(24):1926-1931, 1994.
[4] C.L.Meinert, "Clinical trials: design, conduct and analysis", New York [etc.]: Oxford University Press, 1986. ISBN 0-19-503568-2.
[5] G.Lawrence, "Case record form design", In: Clinical Data Management. Rondel RK., Varley SA., Webb C. (Eds.). New York. John Wiley & Sons Ltd. 1993. Chapter 9. pp. 133-152. ISBN 0-471-94092-5.
[6] W.d'Hollosy, B.Th.Hendriks, W.P.Witjes, F.M.J.Debruyne, H.Wijkstra, "Improvement of data management in scientific urological research.", Proc. IEEE-EMBS, vol. 16, pp. 1394-1395, 1994.
[7] A.Newell, "The knowledge level",. Artif.Intel. 18, pp.87-127, 1982.
[8] M.A.Musen, "Dimensions of knowledge sharing and reuse.", Comput.Biomed.Res. 25, pp. 435-467,1992.