*User Centred Networked Health Care*
*A. Moen et al. (Eds.)*
*IOS Press, 2011*

*doi:10.3233/978-1-60750-806-9-170*

# Identifying Patients for Clinical Trials Using Fuzzy Ternary Logic Expressions on HL7 Messages

Raphael W. MAJEED[a], Rainer RÖHRIG[a,1]
[a] *Department of anesthesia and intensive care medicine*
*Justus-Liebig University Giessen (Germany)*

**Abstract.** Identifying eligible patients is one of the most critical parts of any clinical trial. The process of recruiting patients for the third phase of any clinical trial is usually done manually, informing relevant physicians or putting notes on bulletin boards. While most necessary information is already available in electronic hospital information systems, required data still has to be looked up individually. Most university hospitals make use of a dedicated communication server to distribute information from independent information systems, e.g. laboratory information systems, electronic health records, surgery planning systems. Thus, a theoretical model is developed to formally describe inclusion and exclusion criteria for each clinical trial using a fuzzy ternary logic expression. These expressions will then be used to process HL7 messages from a communication server in order to identify eligible patients.

**Keywords.** Clinical trials, patient recruitment, hl7, communication server, fuzzy logic, ternary logic, data warehouse

## 1. Introduction

Conducting clinical trials comes hand in hand with immense efforts and high costs. Delays or even the failure of a study leads to consequences of ethical and economical nature. Most clinical trials require precisely defined collectives, described by eligibility criteria for inclusion and exclusion. Failures of clinical trials are usually due to falling below necessary patient numbers [1]. On one hand, study centers often underestimate the number of patients actually matching the required eligibility criteria, while on the other hand many patients fail to be enrolled due to nescience of medical personnel.

A government funded research project at five universities thus aims to investigate how the recruiting process for clinical trials can be electronically assisted by hospital information systems (HIS) and clinical information systems (CIS). As a matter of course, implementing patient recruitment functionality in HIS strongly depends on the deployed software. The prevalence of the communication standard HL7 (version 2.x) and the nearly nationwide use of communication servers suggest the possibility to create a generic solution to identify possibly eligible patients for clinical trials. Previous attempts perform database queries [2] or require users to interactively import or enter patient information. Aim of this paper is the development of a solution suitable

[1] Corresponding author: rainer.roehrig@chiru.med.uni-giessen.de

to identify eligible patients for clinical trials by listening to communication server messages. The envisaged automatic inclusion of information from a clinical integration server presents a novel approach.

## 2. Methods

An automated recommendation of eligible patients for clinical trials requires first of all a formalization and electronic description of eligibility criteria. Since we aim at a routine use of the recommendation system to be developed, realistic requirements play an important role in the development process. Therefore, our strategy consists of firstly deducing application-oriented requirements for a formal description of eligibility criteria. Subsequently, a computer processible description language is to be developed which conforms to our requirements. Finally, a feasibility study will be conducted to evaluate the suitability of our approach for routine usage.

### *2.1. Requirements for Formally Describing Eligibility Criteria*

Key requirement for a formal description language is its ability to describe the targeted scenario. Hence, the web site ClinicalTrials.gov is used to find all trials currently in phase III (e.g. recruiting patients) which are enrolled in all five German universities participating in the government funded project. The formal description is required to represent most of the eligibility criteria found in these trials.

Not all eligibility criteria are satisfiable with equal precision. Discussions with local experts in medical informatics on decidability of patient information yielded four distinct criteria groups: If a criterion is based on data from master patient records or laboratory results (a), it is completely decidable once the information arrives. If, however, a criterion bases on the existence a diagnosis (b), it is immanent that the patient was previously examined for the diagnosis and the diagnosis was confirmed. The same also holds for medical procedures, prescribed medication. More difficult are eligibility criteria basing on the nonexistence of certain diagnoses, procedures or medication (c). It is forbidden, to conclude from a nonexistent diagnosis that the patient does not have the diagnosis – he just might not have been examined for the diagnosis yet. Finally, some eligibility criteria might completely resist automatic or electronic verification (d), like conditions concerning patients history, future or intimate information. Since electronic verification of the described criteria groups (a)-(d) occurs with different precision, a formal description language for eligibility criteria is primarily required to satisfy criteria of groups (a), (b) and (c).

Applicability of the defined groups is to be determined by having two experts assign these groups to all criteria independently, with a third expert to resolve conflicts.

### *2.2. Approach*

The concept of interfacing with a communication server follows a primarily passive approach, because no additional action by the medical personnel or the patient is needed. In order to solely rely on HL7 messages for deciding patient eligibility, it has to be investigated whether all required information can be delivered unsolicitedly by the communication server and by what means missing information can be acquired from different sources.

### *2.3. Feasibility Study*

To determine whether the developed formal description model is suitable for routine usage, an evaluation of the model for all trials described in section 2.1 needs to be performed. Additionally, a simple implementation of a software prototype will serve as hint, if interfacing with a communication server is possible without modifications to the server and whether eligible patients can be identified using the description model.

## 3. Results

Our search for clinical trials on the web site ClinicalTrials.gov meeting the previously described criteria yielded 11 relevant trials. All of previously described criteria groups (a)-(d) were present in these trials. Table 1 shows example criteria for groups (a)-(d). Results of two experts assigning those groups to all criteria (Cohen's kappa 0.7) are shown in table 2. Assignment conflicts were resolved by a third expert.

**Table 1.** Examples illustrate the four semantic groups used to categorize eligibility criteria.

| Semantic group | | Example criteria |
|---|---|---|
| (a) Completely decidable facts: | - | "Age 3 Months to 30 Years" |
| Master patient record, | - | "Hemoglobin > 10g/dl" |
| laboratory tests, | - | "No overt renal disease" (creatinine < limit) |
| numeric scores | - | "Performance status ECOG ≥ 3" |
| (b) Partially decidable facts: | - | "Medulloblastoma, cerebral PNET or Ependymoma" |
| positively formulated diagnoses, | - | "More than 4 weeks since prior radiotherapy" |
| procedures or medication | - | "treatment with peginterferon alfa-2A" |
| (c) Undecidable facts: | - | "no experimental drugs" |
| negated diagnoses, medication, procedures | - | "no chronic renal disease" |
| (d) No automatic data | - | "Refactory or relapsed disease" |
| processing: | - | „subject unlikely to comply with protocol " |
| Information about past | - | "no allergy or intolerance to study medication" |
| events/history, | - | "no pre-existing illness preventing treatment" |
| Information about the future | - | „no refusal to use effective contraception" |
| | - | "available for long term follow up through treating center" |

**Table 2.** Results of applying criteria groups to clinical trial eligibility criteria by two experts (Cohen's kappa 0.7), with a third expert to resolve conflicts.

| Trial | Total Criteria | Group (a) | Group (b) | Group (c) | Group (d) |
|---|---|---|---|---|---|
| NCT00749723 | 28 | 11 | 7 | 5 | 5 |
| NCT00876031 | 13 | 1 | 5 | 4 | 3 |
| NCT01011738 | 6 | 5 | 1 | 0 | 0 |
| NCT00733343 | 29 | 2 | 9 | 15 | 3 |
| NCT01077232 | 8 | 1 | 1 | 4 | 2 |
| NCT00526318 | 10 | 2 | 4 | 3 | 1 |
| NCT00554502 | 24 | 6 | 3 | 13 | 2 |
| NCT01155193 | 8 | 2 | 3 | 0 | 3 |
| NCT00290667 | 32 | 5 | 2 | 22 | 3 |
| NCT01127750 | 10 | 1 | 1 | 8 | 0 |
| NCT00410631 | 12 | 3 | 5 | 3 | 1 |

### *3.1. Eligibility Criteria as Propositional Calculus Expressions*

All revised clinical trials from ClinicalTrials.gov describe eligibility criteria in a similar way. The patient is eligible to a clinical trial if and only if all items of a bulleted list of

inclusion criteria evaluate to true and none of the exclusion criteria evaluates to true. Some bulleted items contain alternative conditions of which one suffices to satisfy the criteria. Since all of the reviewed trial descriptions follow this scheme, a formal description language shall reflect this property. The presence the logic terms *and*, *or* and *not* suggests a formalization in propositional calculus. In propositional calculus, boolean formulas displaying the previously described properties are said to be in *conjunctive normal form (CNF)*. Therefore, *CNF* formulas are chosen for formally describing eligibility criteria. The availability of advanced algorithms, like satisfiability problem (SAT) solvers [3] provide an additional advantage.

### *3.2. Fuzzy ternary Logic*

Evaluating previously constructed CNF formulas using clinical trial data within the scope of the planned feasibility study led to the conclusion that classical Boolean interpretation of the CNF formulas is insufficient for the recruitment process. Since most information is missing at the beginning of the recruitment process and only logic values true and false are allowed, logic expressions containing missing data simply evaluate to false resulting in an exclusion/rejection of the patient. In general, the nonexistence of a value does not permit a conclusion to a logic value *false*. Allowing a third logic value *unknown*, leads to three-valued logic also known as ternary logic. Ternary logic still allows the basic logic operations *and*, *or*, *not* which are also used by CNF formulas. A ternary CNF formula may now produce the value *unknown*.

Evaluation of the CNF formulas using eligibility criteria also resulted in a second problem: While inclusion and exclusion criteria are usually sharp in the sense of either true or false, medical parameters vary over time and in precision. Since the final decision whether a patient is eligible or not is always performed by a physician, it is desired that patients slightly outside of the eligible range are also identified to allow a physician to perform a more precise examination of critical values. If, for example, a patient's creatinine value exceeds the permissible range slightly, a physician might still examine the patient to determine whether his condition might change. This leads to fuzzy logic, which enables logic expressions to assume values between 0 (false) and 1 (true). Operations and, or, not are also defined as min(A,B), max(A,B) and 1-A.

It is possible to combine both ternary logic and fuzzy logic to overcome the previously described restrictions.

## 4. Discussion

The chosen approach to describe eligibility criteria of clinical trials with fuzzy ternary logic formulas suffices for the purpose of assisting patient recruitment. Results of Table 2 and the prototype implementation suggest that most eligibility criteria can be sufficiently described using fuzzy ternary logic CNF formulas.

One might question, why to choose logic formula in favor of more powerful approaches like the ARDEN syntax. The advantage of logic formulas lays in its low complexity compared to full programming languages according to formal languages and complexity theory. While most problems are decidable or even solvable for logic formulas, the contrary applies to Turing complete languages, for which e.g. the halting problem is known to be unsolvable [4]. In contrast to previous formalizations (recently

reviewed by Weng et al.[5]), the developed model focuses on decidability of patient information and is especially suited for live recruitment using communication servers.

### *4.1. Storage and Caching of Medical Facts*

The communication server is able to provide information about diagnoses, procedures, laboratory results and medication. Still, many eligibility criteria require information from the patient's clinical history. Using the routine clinical information systems to acquire this information is on the one hand not always practical due to quality and performance reasons and on the other hand results in vendor specific implementations. To overcome this limitation, a clinical data warehouse (e.g. project i2b2) might be used to store incoming HL7 messages from the communication server.

### *4.2. Runtime Restrictions for Deployment to Routine Usage*

Since our goal is using the presented technique in a 1200 bed hospital with more than 40000 patient encounters per year, careful attention is needed regarding processing and resource limitations. After full deployment, any diagnosis, procedure, prescribed or documented medication and laboratory result will be forwarded to our clinical trial patient identification software. Therefore, the software is required to process around twenty thousand HL7 messages per day, at peak times around 30 messages per second.

According to the developed model, the recruiting process of one study for one patient terminates if and only if the generated expression evaluates to either true or false. Consequently a recruitment process might be active concurrently for each and every patient and study, which will sum up to several thousand recruitment processes at any time. Thus, a concept is needed to swap out and store recruitment processes, also to enable the software to stop and resume at a later time. As a clinical data warehouse is already needed for storing and caching medical facts, it might also serve to store the state of recruitment processes. Additionally, the number of concurrently active recruitment processes might be reduced by declaring certain facts as *trigger facts* to explicitly start a recruitment process while storing other facts for later evaluation. Since the study results concluded a general feasibility of the presented concept, the prototype is currently being extended for full functionality.

**Acknowledgements**: Funded by the German ministry for education and research (BMBF), Project 01 EZ 0941 X

## References

[1] Campbell MK, Snowdon C, Francis D et al.: Recruitment to randomised trials: strategies for trial enrolment and participation study. The STEPS study. *Health Technol Assess* 2007;11(48)

[2] Thadani SR, Weng C, Bigger JT, Ennever JF, Wajngurt D. Electronic screening improves efficiency in clinical trial recruitment. *Journal of the American Medical Informatics Association* 2009;16(6):869–73.

[3] Schuler R. An algorithm for the satisfiability problem of formulas in conjunctive normal form. *Journal of Algorithms* 2005;54(1):40–4

[4] Turing AM. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 1937;2(1):230.

[5] Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: A literature review. Journal of biomedical informatics 2010;43(3):451–67.