OWL Model of Clinical Trial Eligibility Criteria Compatible With Partially-known Information

Olivier Dameron[1,2], Paolo Besana[1,2], Oussama Zekri[3], Annabel Bourdé[1,2], Anita Burgun[1,2], and Marc Cuggia[1,2]

[1] Université de Rennes1, UMR936, F-35000 Rennes, France
olivier.dameron@univ-rennes1.fr,
[2] INSERM UMR936, F-35000 Rennes, France
[3] Centre Régional de Lutte Contre le Cancer Eugène Marquis, F-35000 Rennes, France

Abstract. Clinical trials are important for patients, for researchers and for companies. One of the major bottlenecks is patient recruitment. This task requires to match a great quantity of information about the patient with numerous eligibility criteria, in a logically-complex combination. Moreover, the patient's information required by some of the eligibility criteria may not be available at the time of pre-screening. In such situations, the classic approach based on negation as failure ignores the distinction between a trial for which patient eligibility should be rejected and trials for which patient eligibility cannot be asserted, which resuls in underestimating recruitment. We propose an OWL design pattern for modeling eligibility criteria based on the open world assumption to address the missing information problem.

Keywords: open world assumption, ontology design pattern, clinical trial eligibility criteria

1 Introduction

A major focus in all clinical trials is the recruitment of patients. Adequate enrollment provides a base for projected participant retention, resulting in evaluative patient data. Identification of eligible patients for clinical trials (from the principal investigator perspective) or identification of clinical trials in which the patient can be enrolled (from the patient perspective) is an essential phase of clinical research and an active area of medical informatics research. The National Cancer Institute identified several barriers that health care professionals claim in regards to clinical trial participation [1]. Among those barriers, lack of awareness of appropriate clinical trials is frequently mentioned.

Automated tools that help perform a systematic screening either of the potential clinical trials for a patient, or of the potential patients for a clinical trial could overcome this barrier [2]. Efforts have been dedicated to provide a uniform access to heterogeneous data from different sources. The Biomedical Translational Research Information System (BTRIS) is being developed at NIH to consolidate clinical research data [3]. It is intended to simplify data access and analysis of data from active clinical trials and to facilitate reuse of existing data to answer new questions. STRIDE [4] is a platform supporting clinical and translational research consisting of a clinical data warehouse, an application development framework for building research data management applications and a biospecimen data management system. The i2b2 framework integrates medical record and clinical research data [5] and SHRINE [6] handles several sources by providing a federated query tool for clinical data repositories. The ObTiMA system relies on OWL and SWRL to perform semantic mediation between heterogeneous data sources [7]. Lezcano et al. propose an architecture based on OWL to represent patients data from archetypes, and on SWRL rules to perform the reasoning [8]. Several other efforts have been dedicated to the formal representation of clinical trials eligibility criteria to support automated reasoning [9]. Weng et al. performed an extensive literature review [10]. Ross et al. conducted a survey of 1,000 criteria randomly selected from ClinicalTrials.gov and found that 80% of them had a significant semantic complexity [11], with 40% involving some temporal reasoning. Tu et al. proposed an approach to convert free text eligibility criteria into the computable ERGO formalism [12]. O'Connor et al. developed a solution based on OWL and SWRL that supports temporal reasoning and bridges the gap between patients specific data and more general eligibility criteria [13]. The ASTEC (Automatic Selection of clinical Trials based on Eligibility Criteria) project aims at automating the search of prostate cancer clinical trials patients could be enrolled to [14]. It features syntactic and semantic interoperability between the oncologic electronic medical records and the recruitment decision system using a set of international standards (HL7 and NCIT), and the inference method is based on ERGO [15]. The EHR4CR project aims at facilitating clinical trial design and patient recruitment by developing tools and services that reuse data from heterogeneous electronic health records. The TRANSFoRm project has similar objectives for primary care.

All these works on data and criteria representation, integration and reasoning are motivated by the requirement to have the necessary information available at the time of processing the patient's data, and assume that somehow, that will be the case. Missing information that is required for deciding whether a criterion is met leads to underestimating recruitment. Solutions for circumventing this difficulty consist either in making assumptions about the undecided criteria, or in having a pre-screening phase considering a subset of the criteria for which patient's data are assumed to be available. Bayesian belief networks have been used to address the former [16] but require a sensible choice of probability values and may lead to the wrong asumption in particular cases. The latter leaves most of the decision task to human expertise, which provides little added value (if an expert has to handle the difficult criteria, taking the simple pre-screening ones into account adds little to the burden) and is still susceptible to the problem of missing information for the pre-screening criteria.

We propose an OWL design pattern for modeling clinical trial eligibility criteria. This design pattern is based on the open world assumption for handling missing information. It infers whether a patient is eligible or not for a clinical trial, or if no definitive conclusion can be reached.

2 Background

2.1 Modeling eligibility criteria

A clinical trial can be modeled as a pair $< (I_i)_{i=0}^n, (E_j)_{j=0}^m >$ where $(I_i)_{i=0}^n$ is the set of the inclusion criteria, and $(E_j)_{j=0}^m$ is the set of the exclusion criteria. All the eligibility criteria from $(I_i)_{i=0}^n \cup (E_j)_{j=0}^m$ are supposed to be independent from the others (at least in the weak sense: the value of criterion C_k cannot be infered from the combined values of other criteria). Each criterion can be modeled as an unary predicate $C(p)$, where the variable p represents all the information available for the patient. $C(p)$ is true if and only if the criterion is met.

A patient is deemed eligible for a clinical trial if $\underline{\text{all}}$ the inclusion criteria and $\underline{\text{none}}$ of the exclusion criteria are met.

$$\boxed{\text{patient eligible} \Leftrightarrow \bigwedge_{i=0}^{n} I_i(p) \wedge \neg(\bigvee_{j=0}^{m} E_j(p))} \tag{1}$$

Before the final decision on the list of clinical trials a patient is eligible for, there are intermediate pre-screening phases where only the main eligibility criteria of each clinical trial are considered. Such pre-screening sessions rely on subsets of $(I_i)_{i=0}^n$ and $(E_j)_{j=0}^m$, but the decision process remains the same.

For the sake of clarity, in addition to the general case, we will consider a simple clinical trial with two inclusion criteria I_0 and I_1, and two exclusion criteria E_0 and E_1.

$$\text{patient eligible} \Leftrightarrow I_0(p) \wedge I_1(p) \wedge \neg(E_0(p) \vee E_1(p)) \tag{2}$$

For example, these criteria could be:

- I_0: evidence of a prostate adenocarcinoma;
- I_1: absence of metastasis;
- E_0: patient older than 70 years old;
- E_1: evidence of diabetes.

According to equation 2, a patient would be eligible for the clinical trial if and only if he has a prostate adenocarcinoma and has no metastasis and is neither older than 70 years old nor has diabetes.

Because of De Morgan's laws, equation 1 is equivalent to:

$$\boxed{\text{patient eligible} \Leftrightarrow (\bigwedge_{i=0}^{n} I_i(p)) \wedge (\bigwedge_{j=0}^{m} \neg E_j(p))} \tag{3}$$

Even though equation 1 and equation 3 are logically equivalent, the latter is often preferred because it is an uniform conjunction of criteria. Note that the negations in front of the exclusion criteria are purely formal, as both inclusion and exclusion criteria can represent an asserted presence (e.g. prostate adenocarcinoma for I_0 or of diabetes for E_1) or an asserted absence (e.g. metastasis for I_1).

For our example:

$$\text{patient eligible} \Leftrightarrow I_0(p) \wedge I_1(p) \wedge (\neg E_0(p)) \wedge (\neg E_1(p)) \qquad (4)$$

According to equation 3, a patient would be eligible for the clinical trial if and only if he has a prostate adenocarcinoma and has no metastasis and is not older than 70 years old and has not diabetes.

2.2 The problem of unknown Information

Distinction between the patients that we know are not eligible and those that we do not know if they are eligible When a part of the information necessary for determining if at least one criterion is met is unknown, the conjunction of equation 3 can never be true. This necessarily makes the patient not eligible for the clinical trial, whereas the correct interpretation of the situation is that the patient cannot be proven to be eligible. This is different from proving that the patient is not eligible, and indeed, in reality the patient can sometimes be included by assuming the missing values (cf. next section).

For our fictitious clinical trial, we consider a population of nine patients covering all the combinations of "*True*", "*False*" or "*Unknown*" for the inclusion criterion I_1 and the exclusion criterion E_1. Table 1 presents the value of equation 4 and correct inclusion decision for the nine combinations. Among the five patients (p_2, p_5, p_6, p_7 and p_8) for which at least a part of the information is unknown, three (p_2, p_7 and p_8) illustrate a conflict between the value of equation 4 and expected inclusion decision. A strict interpretation of equation 4 leads to the exclusion of the eight patients:

- for three of them (p_0, p_3 and p_4), all the information is available;
- for two of them (p_5 and p_6), some information is unknown, but the available information is sufficient to conclude that the patients are not eligible;
- for the three others (p_2, p_7 and p_8), however, the cause of rejection is either because one of the inclusion criteria cannot be proven (I_1 for p_7 and p_8) or because one of the exclusion criteria cannot be proven to be false (E_1 for p_2 and p_8).

Therefore, if we generalize, equation 3 alone is not enough in the case of partially-known information to make the distinction between the patients we know are not eligible (the first two categories, so this also includes patients for whom a part of the information is unknown) and those we do not know if they are eligible (the third category). This is a problem because patients from the first two categories should be excluded from the clinical trial, whereas those from the third category should be considered for inclusion.

Patient	I_0	I_1	E_0	E_1	$I_0 \wedge I_1 \wedge \neg E_0 \wedge \neg E_1$	Decision
p_0	T	T	F	T	F	Exclude (E_1)
p_1	T	T	F	F	T	Include
p_2	T	T	F	?	**F** cannot assert $\neg E_1$	**Propose** (assume $\neg E_1$)
p_3	T	F	F	T	F	Exclude (both $\neg I_1$ and E_1)
p_4	T	F	F	F	F	Exclude ($\neg I_1$)
p_5	T	F	F	?	F	Exclude ($\neg I_1$)
p_6	T	?	F	T	F	Exclude (E_1)
p_7	T	?	F	F	**F** cannot assert I_1	**Propose** (assume I_1)
p_8	T	?	F	?	**F** cannot assert I_1 cannot assert $\neg E_1$	**Propose** (assume both I_1 and $\neg E_1$)

Table 1. Evaluation of equation 4 and correct inclusion decision for all the possible values of I_1 and E_1, with possibly unknown information

Assuming values for criteria Currently, the case of each patient diagnosed with cancer is examined in a multidisciplinary meeting (MDM), gathering (oncologists, pathologists, surgeons,...). The goal is to determine collectively the best therapeutic strategy for the patient, including consideration of potential inclusion into clinical trials. This preliminary stage is called pre-screening because it takes place before obtaining informed consent (i.e., before enrollment). It mainly relies on retrospective data coming from the patient health record. At this point, all the information necessary for determining the status of each inclusion and exclusion criteria may not be available, but the rationale is to focus on the clinical trials the patient may be eligible for. It should be noted that the missing items may differ between patients. One solution could be to assume the values of the unknown criteria in order to go back to a situation where inclusion or exclusion could be computed using equation 3.

In this case:

- inclusion criteria for which the available information is not sufficient to compute the status are considered to be met;
- exclusion criteria for which the available information is not sufficient to compute the status are considered not to be met.

Therefore, in the case where the available information is not sufficient to compute the status of a criterion, a different status is assumed depending on whether the criterion determines inclusion or exclusion.

Referring to our fictitious clinical trial, the lack of information about the absence of metastasis would lead to the assumption that I_1 is true, whereas the lack of information about diabetes would lead to the assumption that E_1 is false.

This situation raises several issues:

- a different status is assumed depending on whether the criterion determines inclusion or exclusion;
- the assumed status depends on the nature of the criterion (i.e. inclusion or exclusion) and not on its probability;
- one has to remember that the value for at least a criterion has been assumed in order to qualify the inferred eligibility (adamant for p_0 or p_1 vs "under the assumption that..." for p_2, p_7 and p_8);
- this qualification can be difficult to compute (the status of E_1 is unknown for both p_2 and p_5, but p_5 can be confidently excluded whereas p_2 can be included assuming E_1).

2.3 The extent of the missing information problem

To determine the extent of the missing information problem, we analyzed the 286 prostate cancer cases examined during the weekly urology multidisciplinary meetings at Rennes' university hospital between October 2008 and March 2009. This involved 252 patients: 25 of them were examined during two different MDM, and 5 were examined during three different MDM. Before the MDM, the patient's data are collected in a form with 59 fields. The form supports the distinction between known and unknown values (e.g. for “antecedent of neoplasm”, the possible answer are “yes”, “no”, “not specified”).

Overall, 58.64% of the values were unknown. On average, for each case studied in a MDM, 34.6 fields (among 59) had an unknown value.

All of the 286 cases studied had at least some of the 59 fields with an unknown value. Indeed, the case with the most fields filled still missed 19 of them.

54 fields (91.53% of 59) had a missing value in at least one of the 286 cases. The five fields that were systematically filled were: the patient identifier, the MDM date, the patient's gender, the tumor anatomic site and the primary histological type.

During this period, 4 clinical trials related to prostate cancer running at Rennes Comprehensive Cancer Center were considered during the MDM. Table 2 presents the composition of the clinical trials fields and their proportion of missing information. It shows that for each clinical trial, all the patients had at least one missing field that prevented formula 3 to be true (regardless of the values of the known fields).

	CT1	CT2	CT3	CT4
Nb inclusion fields	15	19	17	10
Nb exclusion fields	10	9	13	11
Nb common fields	3	0	3	3
Missing values	50.06%	61.72%	52.99%	42.99%
Nb patients with all inclusion fields known	0	0	1	1
Nb patients with all exclusion fields known	4	3	0	1
Nb patients with all fields known	0	0	0	0
Nb eligible patients	30	23	7	2

Table 2. Importance of unknown information during pre-screening for the four clinical trials of interest

3 Methods

We propose an OWL design pattern for modeling clinical trial eligibility criteria. We then explain how the reasoning unfolds using the fictitious clinical trial from table 1. We validate our approach by verifying if the inferred outcome corresponds to the expected value from table 1. We evaluate our approach on the four clinical trials related to prostate cancer and the 286 cases mentionned at section 2.3. This allows us to quantify the impact of missing information on inclusion rates, as we have seen that in some cases, even partially-known information can lead to certain rejection.

4 Results

4.1 Eligibility criteria design pattern

- for each criterion, create a class C_i (at this point, we do not care if it is an inclusion or an exclusion criteria, or both) and possibly add a necessary and sufficient definition representing the criterion itself (or use SWRL);
- for each criterion, create a class Not_C_i defined as Not_C_i $\equiv$ Criterion $\sqcap\neg$ C_i.This process can be automated;
- for each clinical trial, create a class Ct_k (placeholder);
- for each clinical trial, create a class Ct_k_include as a subclass of Ct_k with a necessary and sufficient definition representing the conjunction of the inclusion criteria and of the exclusion criteria (cf. equation 3) (Ct_k_include $\equiv \bigsqcap_{i=0}^{n}$ I_i $\sqcap \bigsqcap_{j=0}^{m}$ Not_E_j);
- for each clinical trial, create a class Ct_k_exclude (placeholder) as a subclass of Ct_k;
- for each clinical trial, create a class Ct_k_exclude_at_least_one_exclusion_criterion as a subclass of Ct_k_exclude with a necessary and sufficient definition representing the disjunction of the exclusion criteria (Ct_k_exclude_at_least_one_exclusion_criterion $\equiv \bigsqcup_{j=0}^{m}$ E_j);
- for each clinical trial, create a class Ct_k_exclude_at_least_one_failed_inclusion_criterion as a subclass of Ct_k_exclude with a necessary and sufficient definition representing the disjunction of the negated inclusion criteria (Ct_k_exclude_at_least_one_failed_incl_criterion $\equiv \bigsqcup_{i=0}^{n}$ Not_I_i);
- represent the patient's data with instances (Fig. 1 and 2). For the sake of simplicity, we will make the patient an instance of as many C_i as we know he matches criteria, and as many Not_C_j classes as we know he does not match criteria, even if this is ontologically questionable (a patient is not an instance of a criterion). How the patient's data are reconciled with the criteria by making the patient an instance of the criteria is not specified here: it can be manually, or automatically with necessary and sufficient definitions or SWRL rules for the C_i and Not_C_j classes.

4.2 Reasoning

If all the required information is available, after classification the patient will be an instance of each C_i or Not_C_i, and therefore will also be instantiated as either Ct_k_include (like p_1 in Fig. 3), Ct_k_exclude_at_least_one_exclusion_criterion or Ct_k_exclude_at_least_one_failed_inclusion_criterion (so at least we are doing as well as the other systems).

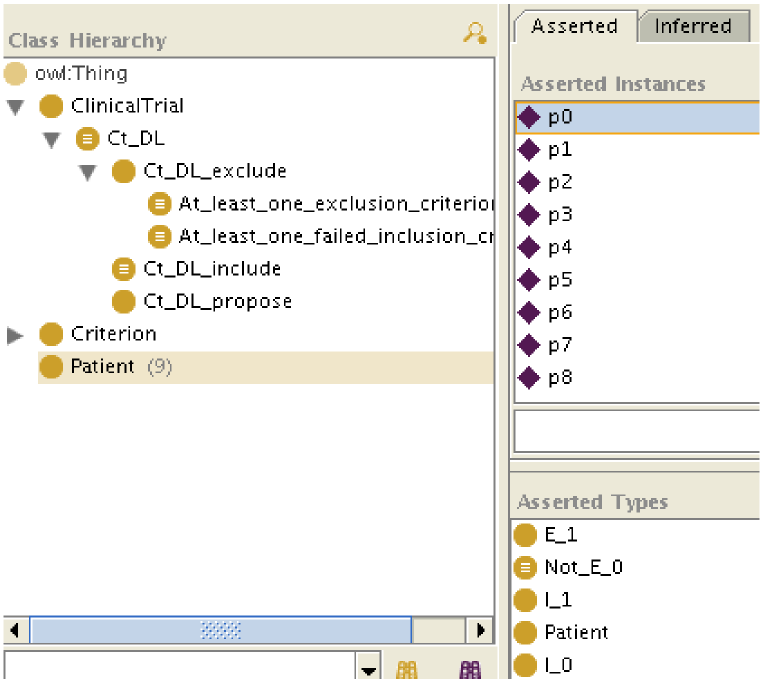


Fig. 1. A patient for who all the information is available

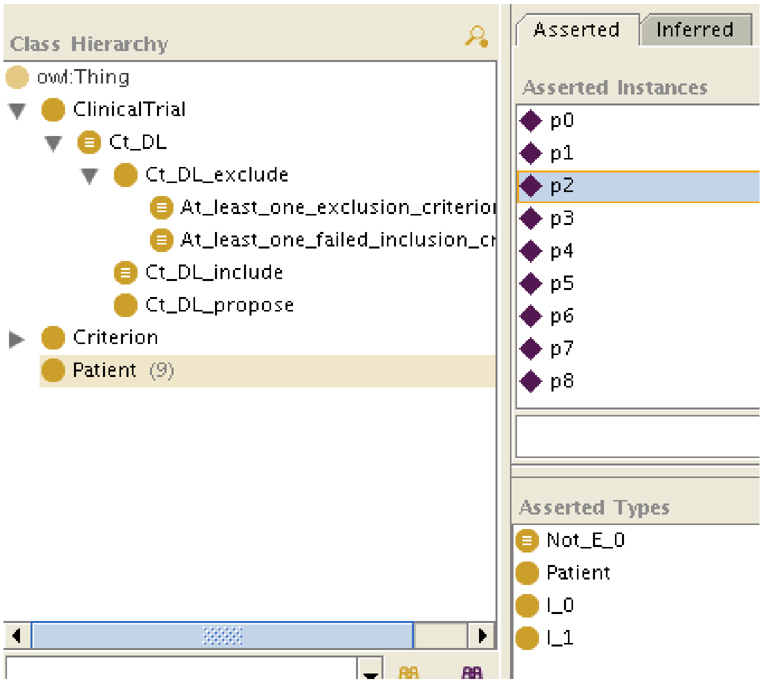


Fig. 2. A patient for who some information is unknown (here about E_1)

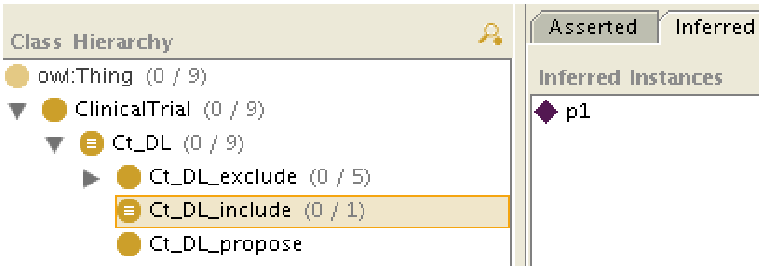
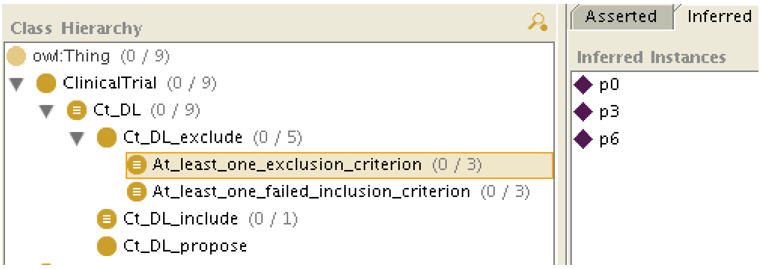


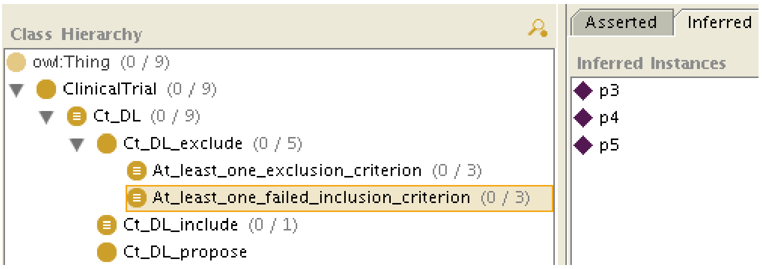
Fig. 3. The class modeling clinical trial inclusion after classification (here patient p_1 can be included)



```
owl:Thing (0 / 9)
▼ ClinicalTrial (0 / 9)
    ▼ Ct_DL (0 / 9)
        ▼ Ct_DL_exclude (0 / 5)
              At_least_one_exclusion_criterion (0 / 3)
              At_least_one_failed_inclusion_criterion (0 / 3)
          Ct_DL_include (0 / 1)
          Ct_DL_propose
```

```
p0
p3
p6
```

Fig. 4. The class modeling clinical trial exclusion because at least one of the exclusion criteria has been met after classification (here patients p_0, p_3 and p_6 match the definition)



```
owl:Thing (0 / 9)
▼ ClinicalTrial (0 / 9)
    ▼ Ct_DL (0 / 9)
        ▼ Ct_DL_exclude (0 / 5)
              At_least_one_exclusion_criterion (0 / 3)
              At_least_one_failed_inclusion_criterion (0 / 3)
          Ct_DL_include (0 / 1)
          Ct_DL_propose
```

```
p3
p4
p5
```

Fig. 5. The class modeling clinical trial exclusion because at least one of the inclusion criteria failed to be met after classification (here patients p_3, p_4 and p_5 match the definition)

If not all the information is available, because of the open world assumption, there will be some criteria for which the patient will neither be classified as an instance of `C_i` nor of `Not_C_i` (Fig. 2), so he will not be classified as an instance of `Ct_k_include` either. However, the patient may be classified as an instance of `Ct_k_exclude_at_least_one_exclusion_criterion` or of `Ct_k_exclude_at_least_one_failed_inclusion_criterion`. As both are subclasses of `Ct_k_exclude`, we will conclude that the patient is not eligible for the clinical trial. We will even know if it is because he matched an exclusion criterion (like p_0, p_3 and p_6 in Fig. 4), because he failed to match an inclusion criterion (like p_3, p_4 and p_5 in Fig. 5), or both (like p_3).

If the patient is neither classified as an instance of `Ct_k_include` nor of `Ct_k_exclude` (or its subclasses), then we will conclude that the patient can be considered for the clinical trial, assuming the missing information will not prevent it (like p_2, p_7 and p_8, who do not appear in Figs. 3, 4 and 5, consistently with Table 1). By retrieving the criteria for which the patient is neither an instance of `C_i` nor of `Not_C_i`, we will know which information is missing.

4.3 Validation

We modeled our fictitious clinical trial from section 2.1 as well as the nine combinations of values from section 2.2[4]. All the results were identical to the decision of table 1.

4.4 Evaluation

We evaluated our model on the first clinical trial (work is ongoing on the three others)[5]. Among the 286 cases, 0 were formally eligible, 122 were potentially eligible, and 164 were not eligible. The 30 cases that were identified as eligible by the experts during the multidisciplinary meetings were all among the 122 proposed by our system (precision was 0.24; recall was 1.0).

It should be noted that the *a posteriori* analysis of the 92 cases proposed by our model but not by the MDM revealed that several were not proposed even if they formally met the eligibility criteria because their Gleason score was deemed too low. We added an inclusion criterion requiring patients to have a Gleason score superior or equal to 7. This resulted in 54 cases potentially eligible, among which were 25 of the 30 actually eligible (precision was 0.46; recall was 0.83). The five false negative cases had a Gleason score of 6. Among the 29 false positive, at least 15 were rejected during the MDM because of additional information not available at the time of pre-screening: 8 because new results indicated that they did not have cancer, 3 because too much information was missing and 4 because other elements such as a relatively young age resulted in proposing a surgical treatment instead of the clinical trial.

[4] `http://www.u936.univ-rennes1.fr/dameron/clinicalTrial/ct-validation.tgz`

[5] `http://www.u936.univ-rennes1.fr/dameron/clinicalTrial/ct-getug14.tgz`

5 Discussion

The analysis of the first clinical trial demonstrates that missing information would have lead to the rejection of all the 30 patients proposed as eligible by the experts during the multidisciplinary meetings. Our approach correctly identified these 30 cases among the 122 it proposed as potentially eligible. This shows that our system confidently rejects non-eligible cases, which leaves more time to examine the others during the multidisciplinary meetings. Moreover, precision can be significatively improved by adding pragmatic criteria that further discriminate the patients who would not be considered as eligible even if they meet the pre-screening criteria. Note that this second step can be kept separated from the formal determination of eligibility but is useful both for the acceptance of the system by the experts and for maintaining the efficiency of the multidisciplinary meetings.

Missing information can partially be handled even with reasoning based on negation as failure using *ad hoc* conversion between inclusion and exclusion criteria. For example, the inclusion criterion "*absence of ischemic heart disease*" can be converted into the exclusion criterion "*presence of ischemic heart disease*". The former will probably never be met because a patient's record only mentions ischemic heart disease when they are present, whereas the latter will (correctly) only exclude those patients having evidence of ischemic heart disease. The problem is that if "*absence of ischemic heart disease*" had been an exclusion criterion, it would likewise have been converted into the inclusion criterion "*presence of ischemic heart disease*" and the system would have (incorrectly, at least during pre-screening) rejected patients whose record does not mention the presence nor the absence ischemic heart disease. Moreover, a criterion can be an inclusion criterion for a clinical trial and an exclusion criterion for another trial, so this strategy is not a general solution to the problem of missing information.

Reasoning about the conjunction of the eligibility criteria should be handled by OWL, which supports the open world assumption, rather than by related technologies such as SWRL which do not. It would be possible to write a SWRL rule that represents the conjunction of criteria (cf. formula 3). However, it is impossible to distinguish situations where we know that one criterion is not met from those where we cannot determine if it is met, because in both cases the rule fill not fire.

Potential applications of our approach are not limited to clinical trials [16]. They cover all clinical decision situations where some information may be missing. We are currently adapting this approach for the determination of pacemaker alerts severity [17]. Electronic health records and clinical reports have been shown to exhibit large amounts of redundant information [18, 19], but Pakhomov et al. observed a discordance between patient-reported symptoms and their (lack of) documentation in the electronic medical records [20]. They noted that this has important implications for research studies that rely on symptom information for patient identification and may have clinical implications that must be evaluated for potential impact on quality of care, patient safety, and outcomes.

6 Conclusion

We showed that ignoring the missing information problem for automatic determination of clinical trial eligibility lead to over-estimate rejection. Systems based on negation as failure infer that the patient is not eligible if it cannot be proved that the is eligible, whereas the situations where it cannot be determined that the patient is eligible nor that he is not eligible should be identified and treated separately. A retrospective analysis of 252 patients with prostate cancer showed that for the four clinical trials of interest, all the patients had at least one missing value that had them rejected whereas 62 of them were actually eligible for at least one of the clinical trials.

We proposed a modeling strategy of eligibility criteria in OWL that leveraged the open world assumption to address the missing information problem. Our approach was able to distinguish a clinical trial for which the patient is eligible, a clinical trial for which we know that the patient is not eligible and a clinical trial for which the patient may be eligible provided that further pieces of information (which we can identify) can be obtained.

Acknowledgments

The ASTEC project is funded by the French Agence Nationale pour la Recherche (ANR 08-TECS-002).

References

1. NCI: Barriers to clinical trial participation. http://www.cancer.gov/clinicaltrials/learningabout/in-depth-program/page7.
2. Marc Cuggia, Paolo Besana, and David Glasspool. Comparing semi-automatic systems for recruitment of patients to clinical trials. International journal of medical informatics, 80(6):371–388, 2011.
3. James J Cimino and Elaine J Ayres. The clinical research data repository of the us national institutes of health. Studies in health technology and informatics, 160(Pt 2):1299–1303, 2010.
4. Henry J Lowe, Todd A Ferris, Penni M Hernandez, and Susan C Weber. Stride–an integrated standards-based translational research informatics platform. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2009:391–395, 2009.
5. Shawn N Murphy, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C Chueh, Susanne Churchill, and Isaac Kohane. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association : JAMIA, 17(2):124–130, 2010.
6. Griffin M Weber, Shawn N Murphy, Andrew J McMurry, Douglas Macfadden, Daniel J Nigrin, Susanne Churchill, and Isaac S Kohane. The shared health research information network (shrine): a prototype federated query tool for clinical data repositories. Journal of the American Medical Informatics Association : JAMIA, 16(5):624–630, 2009.

7. Holger Stenzhorn, Gabriele Weiler, Mathias Brochhausen, Fatima Schera, Vangelis Kritsotakis, Manolis Tsiknakis, Stephan Kiefer, and Norbert Graf. The ObTiMA system - ontology-based managing of clinical trials. Studies in health technology and informatics, 160(Pt 2):1090–1094, 2010.
8. Leonardo Lezcano, Miguel-Angel Sicilia, and Carlos Rodríguez-Solano. Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules. Journal of biomedical informatics, 44(2):343–353, 2010.
9. Ida Sim, Ben Olasov, and Simona Carini. An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method. Journal of Biomedical Informatics, 37(2):108–119, 2004.
10. Chunhua Weng, Samson W Tu, Ida Sim, and Rachel Richesson. Formal representation of eligibility criteria: a literature review. Journal of biomedical informatics, 43(3):451–467, 2009.
11. Jessica Ross, Samson Tu, Simona Carini, and Ida Sim. Analysis of eligibility criteria complexity in clinical trials. AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science, 2010:46–50, 2010.
12. Samson W Tu, Mor Peleg, Simona Carini, Michael Bobak, Jessica Ross, Daniel Rubin, and Ida Sim. A practical method for transforming free-text eligibility criteria into computable criteria. Journal of biomedical informatics, 44(2):239–250, 2010.
13. Martin J O'Connor, Ravi D Shankar, David B Parrish, and Amar K Das. Knowledge-data integration for temporal reasoning in a clinical trial system. International journal of medical informatics, 78 Suppl 1:S77–S85, 2008.
14. Marc Cuggia, Jean-Charles Dufour, Paolo Besana, Olivier Dameron, Regis Duvauferrier, Dominique Fieschi, Catherine Bohec, Annabel Bourdé, Laurent Charlois, Cyril Garde, Isabelle Gibaud, Jean-Francois Laurent, Oussama Zekri, and Marius Fieschi. ASTEC: A system for automatic selection of clinical trials. In Proceedings of the American Medical Informatics Association Conference AMIA, 2011.
15. Paolo Besana, Marc Cuggia, Oussama Zekri, Annabel Bourdé, and Anita Burgun. Using semantic web technologies for clinical trial recruitment. In 9th International Semantic Web Conference (ISWC2010), 2010.
16. L Ohno-Machado, E Parra, S B Henry, S W Tu, and M A Musen. AIDS2: a decision-support tool for decreasing physicians' uncertainty regarding patient eligibility for HIV treatment protocols. Proceedings Symposium on Computer Applications in Medical Care, pages 429–433, 1993.
17. Olivier Dameron, Pascal van Hille, Lynda Temal, Arnaud Rosier, Louise Deléger, Cyril Grouin, Pierre Zweigenbaum, and Anita Burgun. Comparison of OWL and SWRL-based ontology modeling strategies for the determination of pacemaker alerts severity. In Proceedings of the American Medical Informatics Association Conference AMIA, 2011.
18. Jesse O Wrenn, Daniel M Stein, Suzanne Bakken, and Peter D Stetson. Quantifying clinical narrative redundancy in an electronic health record. Journal of the American Medical Informatics Association : JAMIA, 17(1):49–53, 2010.
19. Rui Zhang, Serguei Pakhomov, Bridget T McInnes, and Genevieve B Melton. Evaluating measures of redundancy in clinical texts. AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2011:1612–1620, 2011.
20. Serguei V Pakhomov, Steven J Jacobsen, Christopher G Chute, and Veronique L Roger. Agreement between patient-reported symptoms and their documentation in the medical record. The American journal of managed care, 14(8):530–539, 2008.