Patterns of Clinical Trial Eligibility Criteria

Krystyna Milian[1], Annette ten Teije[1], Anca Bucur[2], Frank van Harmelen[1]
[1] Vrije Universiteit Amsterdam, The Netherlands; [2]Philips Research, The Netherlands

ABSTRACT

Medical research would benefit from automatic methods that support eligibility evaluation for patient enrollment in clinical trials and design of eligibility criteria. In this study we addressed the problem of formalizing eligibility criteria. By analyzing a large set of breast cancer clinical trials we derived a set of patterns, that capture typical structure of conditions, pertaining to syntax and semantics. We qualitatively analyzed their expressivity and evaluated coverage using regular expressions, running experiments on a few thousands of clinical trials also related to other diseases. Based on an early evaluation we conclude that derived patterns cover the language of eligibility criteria to a large extent and may serve as a semi-formal representation. We expect that extending the presented method for pattern recognition with recognition of ontology concepts will facilitate generating computable queries and automated reasoning for various applications.

CLASSIFICATION AND EXAMPLES OF PATTERNS

Based on our experiments, we identified that the following properties would improve the automatic reasoning capabilities with trial eligibility criteria:

- **Time independent status**: present, absent, conditional, potential, not selective
- **Temporal status**: historical, current, planned
- **Specification type**: time frame, including and excluding findings or therapies, value restrictions, purpose of a drug/treatment, co-occurrences, number of occurrences, confirmation, outcome constraint.
- **Medical content**: demographic data (age, gender), clinical data (pregnancy and nursing, menopausal status, adverse reactions), pathology and molecular data, interventions (prior and current therapies).
- **Data source of medical content**. This dimension is dependent on concrete EHR, it is meant to support automatic information extraction.
- **Variability and controllability**: stable, variable, controllable, subjective.
- **Subject:** candidate, family of a candidate

Dimension	%	Example of a pattern	Example of an instance
Time independent status	58		
Absent	16	No concurrent ()	No concurrent endocrine therapy.
Conditional	13	() allowed if ()	Multicentric breast tumors are allowed if all foci are ER-negative.
Temporal status			
Historical	26	No history of ()	No history of brain metastases
Current		Allergy to ()	Allergy to bisphosphonates.
Specification type	72		
Time frame	15	At least () since prior ()	At least 3 weeks since prior steroids.
Exclusions	5	No prior () except for ()	No prior malignancy, except for adequately treated basal cell.
Confirmation	8	confirmed by ()	No metastasis to brain (confirmed by CT or MRI)
Medical content	45		
Clinical	8	No pregnant	Negative pregnancy test
Pathology	7	T () stage	T1-3
Variability and controllability			
Stable	26	Gene () mutation	Known carrier of BRCA1 or BRCA2 mutation
Controllable	1	Must use contraception.	Patients must use effective non-hormonal contraception.
Subject			
Family of a candidate	5	family history of ()	Family history of colon cancer.

PATTERNS RECOGNITION

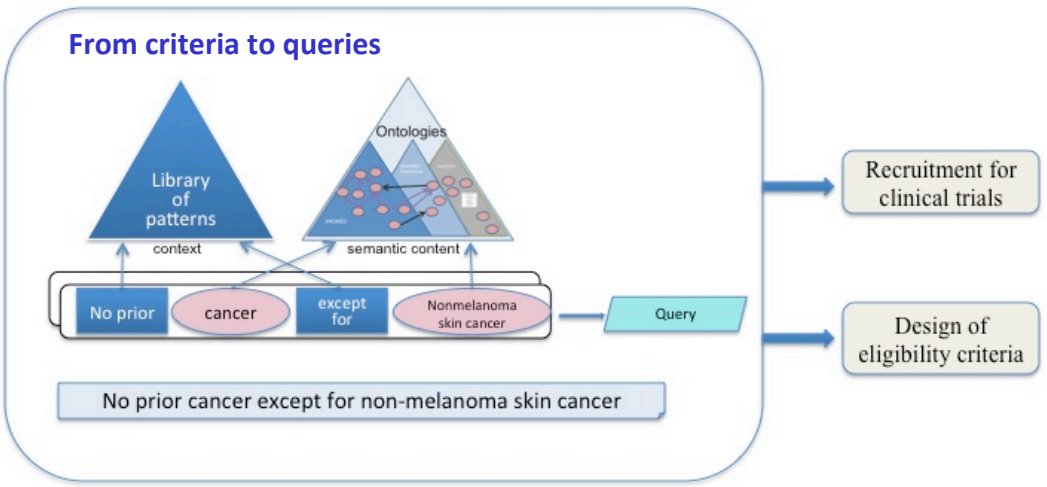
In order to evaluate the coverage of the defined set of patterns across medical domains we analyzed eligibility criteria from breast cancer, lung cancer and diabetes, published at ClinicalTrials.gov. We calculated a number of occurrences of each pattern in the set of eligibility criteria using 342 regular expressions.

	Breast cancer	Lung cancer	Diabetes
No. Trials	3905	2949	5499
No. of sentences processed	111334	119547	86526
Sentences with at least 1 identified patter	71 %	69 %	54%

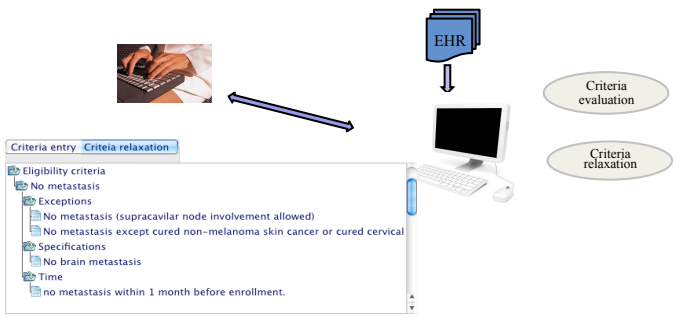
APPLICATIONS

We will use the patterns in the process of formalizing eligibility criteria. Based on annotations of criteria content with the patterns and ontology concepts, we can start generating computable queries.

From criteria to queries



Supporting design of eligibility criteria



SUMMARY

- ✓ We have investigated the possibility of capturing and formalizing the jargon of clinical trial eligibility criteria.
- ✓ We approached the problem by defining a set of 130 patterns that differ in the complexity level. We used 342 regular expressions to identify the patterns in eligibility criteria from breast cancer, lung cancer and diabetes clinical trials and were able to find at least one pattern in 71%, 69% and 54% of lines, respectively.
- ✓ We obtained a method for automatic classification of eligibility criteria according to fine-grained dimensions.

CONCLUSION

- ✓ Our findings indicate that the language used for expressing eligibility criteria is regular enough to be captured to a big extent by the set of defined patterns.